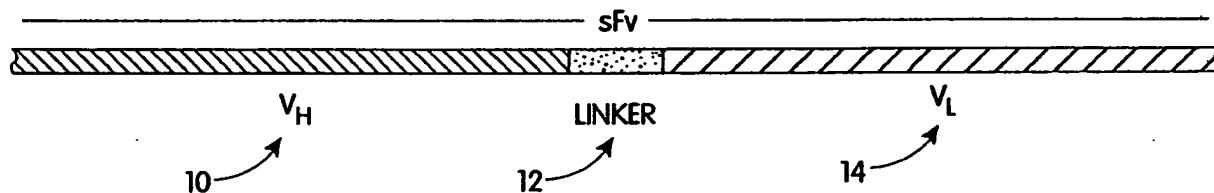




## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>5</sup> :</b> <b>C12N 15/13, 15/62, C12P 21/08</b> <b>C07K 13/00, 15/28, C12N 1/21</b> <b>G01N 33/68, 33/574</b>	<b>A2</b>	<b>(11) International Publication Number:</b> <b>WO 93/16185</b>  <b>(43) International Publication Date:</b> <b>19 August 1993 (19.08.93)</b>
<b>(21) International Application Number:</b> <b>PCT/US93/01055</b> <b>(22) International Filing Date:</b> <b>5 February 1993 (05.02.93)</b>  <b>(30) Priority data:</b> <b>831,967</b> <b>6 February 1992 (06.02.92)</b> <b>US</b>  <b>(71) Applicants:</b> <b>CREATIVE BIOMOLECULES, INC. [US/US]; 35 South Street, Hopkinton, MA 01748 (US). CETUS ONCOLOGY CORPORATION [US/US]; 4560 Horton Street, Emeryville, CA 94608-2916 (US).</b>  <b>(72) Inventors:</b> <b>HUSTON, James, S. ; Five Drew Road, Chestnut Hill, MA 02167 (US). HOUSTON, L., L. ; 1901 Oakcrest Drive, Oakland, CA 94602 (US). RING, David, B. ; 1248 Truman Street, Redwood City, CA 94061 (US). OPPERMANN, Hermann ; 23 Summer Hill Road, Medway, MA 02053 (US).</b>	<b>(74) Agent:</b> <b>KELLEY, Robin, D.; Testa, Hurwitz &amp; Thibault, Exchange Place, 53 State Street, Boston, MA 02109 (US).</b>  <b>(81) Designated States:</b> <b>AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).</b>  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>	

**(54) Title:** BIOSYNTHETIC BINDING PROTEIN FOR CANCER MARKER



**(57) Abstract**

Disclosed is a single-chain Fv (sFv) polypeptide defining a binding site which exhibits the immunological binding properties of an immunoglobulin molecule which binds c-erbB-2 or a c-erbB-2-related tumor antigen, the sFv includes at least two polypeptide domains connected by a polypeptide linker spanning the distance between the C-terminus of one domain and the N-terminus of the other, the amino acid sequence of each of the polypeptide domains includes a set of complementarity determining regions (CDRs) interposed between a set of framework regions (FRs), the CDRs conferring immunological binding to the c-erbB-2 or c-erbB-2-related tumor antigen.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NL	Netherlands
BE	Belgium	GN	Guinea	NO	Norway
BF	Burkina Faso	GR	Greece	NZ	New Zealand
BG	Bulgaria	HU	Hungary	PL	Poland
BJ	Benin	IE	Ireland	PT	Portugal
BR	Brazil	IT	Italy	RO	Romania
CA	Canada	JP	Japan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo	KR	Republic of Korea	SE	Sweden
CH	Switzerland	KZ	Kazakhstan	SK	Slovak Republic
CI	Côte d'Ivoire	LI	Liechtenstein	SN	Senegal
CM	Cameroon	LK	Sri Lanka	SU	Soviet Union
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	MC	Monaco	TG	Togo
DE	Germany	MG	Madagascar	UA	Ukraine
DK	Denmark	ML	Mali	US	United States of America
ES	Spain	MN	Mongolia	VN	Viet Nam
FI	Finland				

### BIOSYNTHETIC BINDING PROTEIN FOR CANCER MARKER

This invention relates in general to novel biosynthetic compositions of matter and, specifically, to biosynthetic antibody binding site (BABS) proteins, and conjugates thereof. Compositions of the invention are useful, for example, in drug and toxin targeting, imaging, immunological treatment of various cancers, and in specific binding assays, affinity purification schemes, and biocatalysis.

#### Background of the Invention

Carcinoma of the breast is the most common malignancy among women in North America, with 130,000 new cases in 1987. Approximately one in 11 women develop breast cancer in their lifetimes, causing this malignancy to be the second leading cause of cancer death among women in the United States, after lung cancer. Although the majority of women with breast cancer present with completely resectable disease, metastatic disease remains a formidable obstacle to cure. The use of adjuvant chemotherapy or hormonal therapy has definite positive impact on disease-free survival and overall survival in selected subsets of women with completely resected primary breast cancer, but a substantial proportion of women still relapse with metastatic disease (see, e.g., Fisher et al. (1986) J. Clin. Oncol. 4:929-941; "The Scottish trial", Lancet (1987) 2:171-175). In spite of the regularly induced objective responses induced by chemotherapy and hormonal therapy in appropriately selected patients, cure of metastatic breast cancer has not been achieved (see e.g., Aisner, et al. (187) J. Clin. Oncol.

- 2 -

5:1523-1533). To this end, many innovative treatment programs including the use of new agents, combinations of agents, high dose therapy (Henderson, ibid.) and increased dose intensity (Kernan et al. (1988) Clin.

5 Invest. 259:3154-3157) have been assembled. Although improvements have been observed, routine achievement of complete remissions of metastatic disease, the first step toward cure, has not occurred. There remains a pressing need for new approaches to treatment.

10 The Fv fragment of an immunoglobulin molecule from IgM, and on rare occasions IgG or IgA, is produced by proteolytic cleavage and includes a non-covalent  $V_H$ - $V_L$  heterodimer representing an intact antigen binding site. A single chain Fv (sFv) polypeptide is a  
15 covalently linked  $V_H$ - $V_L$  heterodimer which is expressed from a gene fusion including  $V_H$ - and  $V_L$ -encoding genes connected by a peptide-encoding linker. See Huston et al., 1988, Proc. Nat. Aca. Sci. 85: 5879, hereby incorporated by reference.

20 U.S. Patent 4,753,894 discloses murine monoclonal antibodies which bind selectively to human breast cancer cells and, when conjugated to ricin A chain, exhibit a TCID 50% against at least one of MCF-7, CAMA-1, SKBR-3, or BT-20 cells of less than about 10 nM.

25 The SKBR-3 cell line is recognized specifically by the monoclonal antibody 520C9. The antibody designated 520C9 is secreted by a murine hybridoma and is now known to recognize c-erbB-2 (Ring et al., 1991, Molecular Immunology 28:915).

- 3 -

Summary of the Invention

The invention features the synthesis of a class of novel proteins known as single chain Fv (sFv) polypeptides, which include biosynthetic single  
5 polypeptide chain binding sites (BABS) and define a binding site which exhibits the immunological binding properties of an immunoglobulin molecule which binds c-erbB-2 or a c-erbB-2-related tumor antigen.

The sFv includes at least two polypeptide domains  
10 connected by a polypeptide linker spanning the distance between the carboxy (C)- terminus of one domain and the amino (N)- terminus of the other domain, the amino acid sequence of each of the polypeptide domains including a set of complementarity determining regions (CDRs)  
15 interposed between a set of framework regions (FRs), the CDRs conferring immunological binding to c-erbB-2 or a c-erbB-2 related tumor antigen.

In its broadest aspects, this invention features single-chain Fv polypeptides including biosynthetic  
20 antibody binding sites, replicable expression vectors prepared by recombinant DNA techniques which include and are capable of expressing DNA sequences encoding these polypeptides, methods for the production of these polypeptides, methods of imaging a tumor expressing  
25 c-erbB-2 or a c-erbB-2-related tumor antigen, and methods of treating a tumor using targetable therapeutic agents by virtue of conjugates or fusions with these polypeptides.

As used herein, the term "immunological binding"  
30 or "immunologically reactive" refers to the non-covalent interactions of the type that occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific; "c-erbB-2" refers to a

- 4 -

protein antigen expressed on the surface of tumor cells, such as breast and ovarian tumor cells, which is an approximately 200,000 molecular weight acidic glycoprotein having an isoelectric point of about 5.3 and including the amino acid sequence set forth in SEQ ID NOS:1 and 2. A "c-erbB-2-related tumor antigen" is a protein located on the surface of tumor cells, such as breast and ovarian tumor cells, which is antigenically related to the c-erbB-2 antigen, i.e., bound by an immunoglobulin that is capable of binding the c-erbB-2 antigen, examples of such immunoglobulins being the 520C9, 741F8, and 454C11 antibodies; or which has an amino acid sequence that is at least 80% homologous, preferably 90% homologous, with the amino acid sequence of c-erbB-2. An example of a c-erbB-2 related antigen is the receptor for epidermal growth factor.

An sFv CDR that is "substantially homologous with" an immunoglobulin CDR retains at least 70%, preferably 80% or 90%, of the amino acid sequence of the immunoglobulin CDR, and also retains the immunological binding properties of the immunoglobulin.

The term "domain" refers to that sequence of a polypeptide that folds into a single globular region in its native conformation, and may exhibit discrete binding or functional properties. The term "CDR" or complementarity determining region, as used herein, refers to amino acid sequences which together define the binding affinity and specificity of the natural Fv region of a native immunoglobulin binding site, or a synthetic polypeptide which mimics this function. CDRs typically are not wholly homologous to hypervariable regions of natural Fvs, but rather may also include specific amino acids or amino acid sequences which

- 5 -

flank the hypervariable region and have heretofore been considered framework not directly determinative of complementarity. The term "FR" or framework region, as used herein, refers to amino acid sequences which are naturally found between CDRs in immunoglobulins.

Single-chain Fv polypeptides produced in accordance with the invention include biosynthetically-produced novel sequences of amino acids defining polypeptides designed to bind with a preselected c-erbB-2 or related antigen material. The structure of these synthetic polypeptides is unlike that of naturally occurring antibodies, fragments thereof, or known synthetic polypeptides or "chimeric antibodies" in that the regions of the single-chain Fv responsible for specificity and affinity of binding (analogous to native antibody variable ( $V_H/V_L$ ) regions) may themselves be chimeric, e.g., include amino acid sequences derived from or homologous with portions of at least two different antibody molecules from the same or different species. These analogous  $V_H$  and  $V_L$  regions are connected from the N-terminus of one to the C-terminus of the other by a peptide bonded biosynthetic linker peptide.

The invention thus provides a single-chain Fv polypeptide defining at least one complete binding site capable of binding c-erbB-2 or a c-erbB-2-related tumor antigen. One complete binding site includes a single contiguous chain of amino acids having two polypeptide domains, e.g.,  $V_H$  and  $V_L$ , connected by a amino acid linker region. An sFv that includes more than one complete binding site capable of binding a c-erbB-2-related antigen, e.g., two binding sites, will be a single contiguous chain of amino acids having four polypeptide domains, each of which is covalently linked

- 6 -

by an amino acid linker region, e.g.,  $V_{H1}$ -linker- $V_{L1}$ -  
linker- $V_{H2}$ -linker- $V_{L2}$ . sFv's of the invention may  
include any number of complete binding sites ( $V_{Hn}$ -  
linker- $V_{Ln}$ )<sub>n</sub>, where  $n > 1$ , and thus may be a single  
5 contiguous chain of amino acids having  $n$  antigen  
binding sites and  $n \times 2$  polypeptide domains.

In one preferred embodiment of the invention, the  
single-chain Fv polypeptide includes CDRs that are  
substantially homologous with at least a portion of the  
10 amino acid sequence of CDRs from a variable region of  
an immunoglobulin molecule from a first species, and  
includes FRs that are substantially homologous with at  
least a portion of the amino acid sequence of FRs from  
a variable region of an immunoglobulin molecule from a  
15 second species. Preferably, the first species is mouse  
and the second species is human.

The amino acid sequence of each of the  
polypeptide domains includes a set of CDRs interposed  
between a set of FRs. As used herein, a "set of CDRs"  
20 refers to 3 CDRs in each domain, and a "set of FRs"  
refers to 4 FRs in each domain. Because of structural  
considerations, an entire set of CDRs from an  
immunoglobulin may be used, but substitutions of  
particular residues may be desirable to improve  
25 biological activity, e.g., based on observations of  
conserved residues within the CDRs of immunoglobulin  
species which bind c-erbB-2 related antigens.

In another preferred aspect of the invention, the  
CDRs of the polypeptide chain have an amino acid  
30 sequence substantially homologous with the CDRs of the  
variable region of any one of the 520C9, 741F8, and  
454C11 monoclonal antibodies. The CDRs of the 520C9  
antibody are set forth in the Sequence Listing as amino  
acid residue numbers 31 through 35, 50 through 66, 99



- 7 -

through 104, 159 through 169, 185 through 191, and 224 through 232 in SEQ ID NOS: 3 and 4, and amino acid residue numbers 31 through 35, 50 through 66, 99 through 104, 157 through 167, 183 through 189, and 222 through 230 in SEQ ID NOS: 5, and 6.

In one embodiment, the sFv is a humanized hybrid molecule which includes CDRs from the mouse 520C9 antibody interposed between FRs derived from one or more human immunoglobulin molecules. This hybrid sFv thus contains binding regions which are highly specific for the c-erbB-2 antigen or c-erbB-2-related antigens held in proper immunochemical binding conformation by human FR amino acid sequences, and thus will be less likely to be recognized as foreign by the human body.

In another embodiment, the polypeptide linker region includes the amino acid sequence set forth in the Sequence Listing as amino acid residue numbers 123 through 137 in SEQ ID NOS:3 and 4, and as amino acid residues 1-16 in SEQ ID NOS:11 and 12. In other embodiments, the linker sequence has the amino acid sequence set forth in the Sequence Listing as amino acid residues 121-135 in SEQ ID NOS:5 and 6, or the amino acid sequence of residues 1-15 in SEQ ID NOS:13 and 14.

The single polypeptide chain described above also may include a remotely detectable moiety bound thereto to permit imaging or radioimmunotherapy of tumors bearing a c-erbB-2 or related tumor antigen. "Remotely detectable" moiety means that the moiety that is bound to the sFv may be detected by means external to and at a distance from the site of the moiety. Preferable remotely detectable moieties for imaging include radioactive atom such as <sup>99m</sup>Tc, a gamma emitter. Preferable nucleotides for high dose

- 8 -

radioimmunotherapy include radioactive atoms such as, ( $^{90}\text{Yttrium}$  ( $^{90}\text{Yt}$ ),  $^{131}\text{Iodine}$  ( $^{131}\text{I}$ ) or  $^{111}\text{Indium}$  ( $^{111}\text{In}$ ).

In addition, the sFv may include a fusion protein  
5 derived from a gene fusion, such that the expressed  
sFv fusion protein includes an ancillary polypeptide  
that is peptide bonded to the binding site polypeptide.  
In some preferred aspects, the ancillary polypeptide  
segment also has a binding affinity for a c-erbB-2 or  
10 related antigen and may include a third and even a  
fourth polypeptide domain, each comprising an amino  
acid sequence defining CDRs interposed between FRs, and  
which together form a second single polypeptide chain  
biosynthetic binding site similar to the first  
15 described above.

In other aspects, the ancillary polypeptide  
sequence forms a toxin linked to the N or C terminus of  
the sFv, e.g., at least a toxic portion of Pseudomonas  
exotoxin, phytolectin, ricin, ricin A chain, or  
20 diphtheria toxin, or other related proteins known as  
ricin A chain-like ribosomal inhibiting proteins, i.e.,  
proteins capable of inhibiting protein synthesis at the  
level of the ribosome, such as pokeweed antiviral  
protein, gelonin, and barley ribosomal protein  
25 inhibitor. In still another aspect, the sFv may  
include at least a second ancillary polypeptide or  
moiety which will promote internalization of the sFv.

The invention also includes a method for  
producing sFv, which includes the steps of providing a  
30 replicable expression vector which includes and which  
expresses a DNA sequence encoding the single  
polypeptide chain; transfecting the expression vector  
into a host cell to produce a transformant; and  
culturing the transformant to produce the sFv  
35 polypeptide.

- 9 -

The invention also includes a method of imaging a tumor expressing a c-erbB-2 or related tumor antigen. This method includes the steps of providing an imaging agent including a single-chain Fv polypeptide as described above, and a remotely detectable moiety linked thereto; administering the imaging agent to an organism harboring the tumor in an amount of the imaging agent with a physiologically-compatible carrier sufficient to permit extracorporeal detection of the tumor; and detecting the location of the moiety in the subject after allowing the agent to bind to the tumor and unbound agent to have cleared sufficiently to permit visualization of the tumor image.

The invention also includes a method of treating cancer by inhibiting in vivo growth of a tumor expressing a c-erbB-2 or related antigen, the method including administering to a cancer patient a tumor inhibiting amount of a therapeutic agent which includes an sFv of the invention and at least a first moiety peptide bonded thereto, and which has the ability to limit the proliferation of a tumor cell.

Preferably, the first moiety includes a toxin or a toxic fragment thereof, e.g., ricin A; or includes a radioisotope sufficiently radioactive to inhibit proliferation of the tumor cell, e.g.,  $^{90}\text{Yt}$ ,  $^{111}\text{In}$ , or  $^{131}\text{I}$ . The therapeutic agent may further include at least a second moiety that improves its effectiveness.

The clinical administration of the single-chain Fv or appropriate sFv fusion proteins of the invention, which display the activity of native, relatively small Fv of the corresponding immunoglobulin, affords a number of advantages over the use of larger fragments or entire antibody molecules. The single chain Fv and sFv fusion proteins of this invention offer fewer

- 10 -

cleavage sites to circulating proteolytic enzymes and thus offer greater stability. They reach their target tissue more rapidly, and are cleared more quickly from the body, which makes them ideal imaging agents for tumor detection and ideal radioimmunotherapeutic agents for tumor killing. They also have reduced non-specific binding and immunogenicity relative to murine immunoglobulins. In addition, their expression from single genes facilitates targeting applications by fusion to other toxin proteins or peptide sequences that allow specific coupling to other molecules or drugs. In addition, some sFv analogues or fusion proteins of the invention have the ability to promote the internalization of c-erbB-2 or related antigens expressed on the surface of tumor cells when they are bound together at the cell surface. These methods permit the selective killing of cells expressing such antigens with the single-chain-Fv-toxin fusion of appropriate design. sFv-toxin fusion proteins of the invention possess 15-200-fold greater tumor cell killing activity than conjugates which include a toxin that is chemically crosslinked to whole antibody or Fab.

Overexpression of c-erbB-2 or related receptors on malignant cells thus allows targeting of sFv species to the tumor cells, whether the tumor is well-localized or metastatic. In the above cases, the internalization of sFv-toxin fusion proteins permits specific destruction of tumor cells bearing the over expressed c-erbB-2 or related antigen. In other cases, depending on the infected cells, the nature of the malignancy, or other factors operating in a given individual, the same c-erbB-2 or related receptors may be poorly internalized or even represent a static tumor antigen

- 11 -

population. In this event, the single-chain Fv and its fusion proteins can also be used productively, but in a different mode than applicable to internalization of the toxin fusion. Where c-erbB-2 receptor/sFv or sFv fusion protein complexes are poorly internalized, toxins, such as ricin A chain, which operate cytoplasmically by inactivation of ribosomes, are not effective to kill cells. Nevertheless, single-chain unfused Fv is useful, e.g., for imaging or radioimmunotherapy, and bispecific single-chain Fv fusion proteins of various designs, i.e., that have two distinct binding sites on the same polypeptide chain, can be used to target via the two antigens for which the molecule is specific. For example, a bispecific single-chain antibody may have specificity for both the c-erbB-2 and CD3 antigens, the latter of which is present on cytotoxic lymphocytes (CTLs). This bispecific molecule could thus mediate antibody dependent cellular cytotoxicity (ADCC) that results in CTL-induced lysis of tumor cells. Similar results could be obtained using a bispecific single-chain Fv specific for c-erbB-2 and the Fcγ receptor type I or II. Other bispecific sFv formulations include domains with c-erbB-2 specificity paired with a growth factor domain specific for hormone or growth factor receptors, such as receptors for transferrin or epidermal growth factor (EGF).

- 12 -

Brief Description of the Drawings

The foregoing and other objects of this invention, the various features thereof, as well as the invention itself, may be more fully understood from the following description, when read together with the accompanying drawings.

FIG. 1A is a schematic drawing of a DNA construct encoding an sFv of the invention, which shows the  $V_H$  and  $V_L$  encoding domains and the linker region; FIG. 1B is a schematic drawing of the structure of Fv illustrating  $V_H$  and  $V_L$  domains, each of which comprises three complementarity determining regions (CDRs) and four framework regions (FRs) for monoclonal 520C9, a well known and characterized murine monoclonal antibody specific for c-erbB-2;

FIGS. 2A-2E are schematic representations of embodiments of the invention, each of which comprises a biosynthetic single-chain Fv polypeptide which recognizes a c-erbB-2-related antigen: FIG. 2A is an sFv having a pendant leader sequence, FIG. 2B is an sFv-toxin (or other ancillary protein) construct, and FIG. 2C is a bivalent or bispecific sFv construct; FIG. 2D is a bivalent sFv having a pendant protein attached to the carboxyl-terminal end; FIG. 2E is a bivalent sFv having pendant proteins attached to both amino- and carboxyl-terminal ends.

FIG. 3 is a diagrammatic representation of the construction of a plasmid encoding the 520C9 sFv-ricin A fused immunotoxin gene; and

FIG. 4 is a graphic representation of the results of a competition assay comparing the c-erbB-2 binding activity of the 520C9 monoclonal antibody (specific for c-erbB-2), an Fab fragment of that monoclonal antibody (filled dots), and different affinity purified

- 13 -

fractions of the single-chain-Fv binding site for  
c-erbB-2 constructed from the variable regions of the  
520C9 monoclonal antibody (sFv whole sample (+), sFv  
bound and eluted from a column of immobilized  
5 extracellular domain of C-erbB-2 (squares) and sFv  
flow-through (unbound, \*)).

- 14 -

Detailed Description of the Invention

Disclosed are single-chain Fv's and sFv fusion proteins having affinity for a c-erbB-2-related antigen expressed at high levels on breast and ovarian cancer cells and on other tumor cells as well, in certain other forms of cancer. The polypeptides are characterized by one or more sequences of amino acids constituting a region which behaves as a biosynthetic antibody binding site. As shown in FIG. 1, the sites comprise heavy chain variable region ( $V_H$ ) 10, light chain variable region ( $V_L$ ) 14 single chains wherein  $V_H$  10 and  $V_L$  14 are attached by polypeptide linker 12. The binding domains include CDRs 2, 4, 6 and 2', 4', 6' from immunoglobulin molecules able to bind a c-erbB-2-related tumor antigen linked to FRs 32, 34, 36, 38 and 32', 34', 36' 38' which may be derived from a separate immunoglobulin. As shown in FIGS. 2A, 2B, and 2C, the BABS single polypeptide chains ( $V_H$  10,  $V_L$  14 and linker 12) may also include remotely detectable moieties and/or other polypeptide sequences 16, 18, or 22, which function e.g., as an enzyme, toxin, binding site, or site of attachment to an immobilization matrix or radioactive atom. Also disclosed are methods for producing the proteins and methods of their use.

The single-chain Fv polypeptides of the invention are biosynthetic in the sense that they are synthesized and recloned in a cellular host made to express a protein encoded by a plasmid which includes genetic sequence based in part on synthetic DNA, that is, a recombinant DNA made from ligation of plural, chemically synthesized and recloned oligonucleotides, or by ligation of fragments of DNA derived from the genome of a hybridoma, mature B cell clone, or a cDNA library derived from such natural sources. The



- 15 -

proteins of the invention are properly characterized as "antibody binding sites" in that these synthetic single polypeptide chains are able to refold into a 3-dimensional conformation designed specifically to have affinity for a preselected c-erbB-2 or related tumor antigen. Single-chain Fv's may be produced as described in PCT application US88/01737, which corresponds to USSN 342,449, filed February 6, 1989, and claims priority from USSN 052,800, filed May 21, 1987, assigned to Creative BioMolecules, Inc., hereby incorporated by reference. The polypeptides of the invention are antibody-like in that their structure is patterned after regions of native antibodies known to be responsible for c-erbB-2-related antigen recognition.

More specifically, the structure of these biosynthetic antibody binding sites (BABS) in the region which imparts the binding properties to the protein, is analogous to the Fv region of a natural antibody to a c-erbB-2 or related antigen. It includes a series of regions consisting of amino acids defining at least three polypeptide segments which together form the tertiary molecular structure responsible for affinity and binding. The CDRs are held in appropriate conformation by polypeptide segments analogous to the framework regions of the Fv fragment of natural antibodies.

The CDR and FR polypeptide segments are designed empirically based on sequence analysis of the Fv region of preexisting antibodies, such as those described in U.S. Patent No. 4,753,894, herein incorporated by reference, or of the DNA encoding such antibody molecules.

- 16 -

One such antibody, 520C9, is a murine monoclonal antibody that is known to react with an antigen expressed by the human breast cancer cell line SK-Br-3 (U.S. Patent 4,753,894). The antigen is an

5 approximately 200 kD acidic glycoprotein that has an isoelectric point of 5.3, and is present at about 5 million copies per cell. The association constant measured using radiolabelled antibody is approximately  $4.6 \times 10^8 \text{ M}^{-1}$ .

10 In one embodiment, the amino acid sequences constituting the FRs of the single polypeptide chains are analogous to the FR sequences of a first preexisting antibody, for example, a human IgG. The amino acid sequences constituting the CDRs are  
15 analogous to the sequences from a second, different preexisting antibody, for example, the CDRs of a rodent or human IgG which recognizes c-erbB-2 or related antigens expressed on the surface of ovarian and breast tumor cells. Alternatively, the CDRs and FRs may be  
20 copied in their entirety from a single preexisting antibody from a cell line which may be unstable or, difficult to culture; e.g., an sFv-producing cell line that is based upon a murine, mouse/human, or human monoclonal antibody-secreting cell line.

25 Practice of the invention enables the design and biosynthesis of various reagents, all of which are characterized by a region having affinity for a preselected c-erbB-2 or related antigen. Other regions of the biosynthetic protein are designed with the  
30 particular planned utility of the protein in mind. Thus, if the reagent is designed for intravascular use in mammals, the FRs may include amino acid sequences that are similar or identical to at least a portion of the FR amino acids of antibodies native to that

- 17 -

mammalian species. On the other hand, the amino acid sequences that include the CDRs may be analogous to a portion of the amino acid sequences from the hypervariable region (and certain flanking amino acids) of an antibody having a known affinity and specificity for a c-erbB-2 or related antigen that is from, e.g., a mouse or rat, or a specific human antibody or immunoglobulin.

Other sections of native immunoglobulin protein structure, e.g., C<sub>H</sub> and C<sub>L</sub>, need not be present and normally are intentionally omitted from the biosynthetic proteins of this invention. However, the single polypeptide chains of the invention may include additional polypeptide regions defining a leader sequence or a second polypeptide chain that is bioactive, e.g., a cytokine, toxin, ligand, hormone, immunoglobulin domain(s), or enzyme, or a site onto which a toxin, drug, or a remotely detectable moiety, e.g., a radionuclide, can be attached.

One useful toxin is ricin, an enzyme from the castor bean that is highly toxic, or the portion of ricin that confers toxicity. At concentrations as low as 1 ng/ml ricin efficiently inhibits the growth of cells in culture. The ricin A chain has a molecular weight of about 30,000 and is glycosylated. The ricin B chain has a larger size (about 34,000 molecular weight) and is also glycosylated. The B chain contains two galactose binding sites, one in each of the two domains in the folded subunit. The crystallographic structure for ricin shows the backbone tracing of the A chain. There is a cleft, which is probably the active site, that runs diagonally across the molecule. Also present is a mixture of  $\alpha$ -helix,  $\beta$ -structure, and irregular structure in the molecule.

- 18 -

The A chain enzymatically inactivates the 60S ribosomal subunit of eucaryotic ribosomes. The B chain binds to galactose-based carbohydrate residues on the surfaces of cells. It appears to be necessary to bind the toxin to the cell surface, and also facilitates and participates in the mechanics of entry of the toxin into the cell. Because all cells have galactose-containing cell surface receptors, ricin inhibits all types of mammalian cells with nearly the same efficiency.

Ricin A chain and ricin B chain are encoded by a gene that specifies both the A and B chains. The polypeptide synthesized from the mRNA transcribed from the gene contains A chain sequences linked to B chain sequences by a 'J' (for joining) peptide. The J peptide fragment is removed by post-translational modification to release the A and B chains. However, A and B chains are still held together by the interchain disulfide bond. The preferred form of ricin is recombinant A chain as it is totally free of B chain and, when expressed in E. coli, is unglycosylated and thus cleared from the blood more slowly than the glycosylated form. The specific activity of the recombinant ricin A chain against ribosomes and that of native A chain isolated from castor bean ricin are equivalent. An amino acid sequence and corresponding nucleic acid sequence of ricin A chain is set forth in the Sequence Listing as SEQ ID NOS:7 and 8.

Recombinant ricin A chain, plant-derived ricin A chain, deglycosylated ricin A chain, or derivatives thereof, can be targeted to a cell expressing a c-erbB-2 or related antigen by the single-chain Fv polypeptide of the present invention. To do this, the sFv may be chemically crosslinked to ricin A chain or

- 19 -

an active analog thereof, or in a preferred embodiment a single-chain Fv-ricin A chain immunotoxin may be formed by fusing the single-chain Fv polypeptide to one or more ricin A chains through the corresponding gene fusion. By replacing the B chain of ricin with an antibody binding site to c-erbB-2 or related antigens, the A chain is guided to such antigens on the cell surface. In this way the selective killing of tumor cells expressing these antigens can be achieved. This selectivity has been demonstrated in many cases against cells grown in culture. It depends on the presence or absence of antigens on the surface of the cells to which the immunotoxin is directed.

The invention includes the use of humanized single-chain-Fv binding sites as part of imaging methods and tumor therapies. The proteins may be administered by intravenous or intramuscular injection. Effective dosages for the single-chain Fv constructs in antitumor therapies or in effective tumor imaging can be determined by routine experimentation, keeping in mind the objective of the treatment.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions or dispersions. In all cases, the form must be sterile and must be fluid so as to be easily administered by syringe. It must be stable under the conditions of manufacture and storage, and must be preserved against the contaminating action of microorganisms. This may, for example, be achieved by filtration through a sterile 0.22 micron filter and/or lyophilization followed by sterilization with a gamma ray source.

Sterile injectable solutions are prepared by incorporating the single chain constructs of the invention in the required amount in the appropriate

- 20 -

solvent, such as sodium phosphate-buffered saline, followed by filter sterilization. As used herein, "a physiologically acceptable carrier" includes any and all solvents, dispersion media, antibacterial and antifungal agents that are non-toxic to humans, and the like. The use of such media and agents for pharmaceutically active substances is well known in the art. The media or agent must be compatible with maintenance of proper conformation of the single polypeptide chains, and its use in the therapeutic compositions. Supplementary active ingredients can also be incorporated into the compositions.

A bispecific single-chain Fv could also be fused to a toxin. For example, a bispecific sFv construct with specificity for c-erbB-2 and the transferrin receptor, a target that is rapidly internalized, would be an effective cytolytic agent due to internalization of the transferrin receptor/sFv-toxin complex. An sFv fusion protein may also include multiple protein domains on the same polypeptide chain, e.g., EGF-sFv-ricin A, where the EGF domain promotes internalization of toxin upon binding of sFv through interaction with the EGF receptor.

The single polypeptide chains of the invention can be labelled with radioisotopes such as Iodine-131, Indium-111, and Technetium-99m, for example. Beta emitters such as Technetium-99m and Indium-111 are preferred because they are detectable with a gamma camera and have favorable half-lives for imaging in vivo. The single polypeptide chains can be labelled, for example, with radioactive atoms and as Yttrium-90, Technetium-99m, or Indium-111 via a conjugated metal chelator (see, e.g., Khaw et al. (1980) Science 209:295; Gansow et al., U.S. Patent No. 4,472,509;

- 21 -

Hnatowich, U.S. Patent No. 4,479,930), or by other standard means of isotope linkage to proteins known to those with skill in the art.

The invention thus provides intact binding sites  
5 for c-erbB-2 or related antigens that are analogous to  $V_H$ - $V_L$  dimers linked by a polypeptide sequence to form a composite  $(V_H$ -linker- $V_L)_n$  or  $(V_L$ -linker- $V_H)_n$  polypeptide, where n is equal to or greater than 1, which is essentially free of the remainder of the  
10 antibody molecule, and which may include a detectable moiety or a third polypeptide sequence linked to each  $V_H$  or  $V_L$ .

FIGs. 2A-2E illustrate examples of protein structures embodying the invention that can be produced  
15 by following the teaching disclosed herein. All are characterized by at least one biosynthetic sFv single chain segment defining a binding site, and containing amino acid sequences including CDRs and FRs, often derived from different immunoglobulins, or sequences  
20 homologous to a portion of CDRs and FRs from different immunoglobulins.

FIG. 2A depicts single polypeptide chain sFv 100 comprising polypeptide 10 having an amino acid sequence analogous to the heavy chain variable region ( $V_H$ ) of a  
25 given anti-c-erbB-2 monoclonal antibody, bound through its carboxyl end to polypeptide linker 12, which in turn is bound to polypeptide 14 having an amino acid sequence analogous to the light chain variable region ( $V_L$ ) of the anti-c-erbB-2 monoclonal. Of course, the  
30 light and heavy chain domains may be in reverse order. Linker 12 should be at least long enough (e.g., about 10 to 15 amino acids or about 40 Angstroms) to permit chains 10 and 14 to assume their proper conformation and interdomain relationship.

- 22 -

Linker 12 may include an amino acid sequence homologous to a sequence identified as "self" by the species into which it will be introduced, if drug use is intended. Unstructured, hydrophilic amino acid sequences are preferred. Such linker sequences are set forth in the Sequence Listing as amino acid residue numbers 116 through 135 in SEQ ID NOS:3, 4, 5, and 6, which include part of the 16 amino acid linker sequences set forth in the Sequence Listing SEQ ID NOS:12 and 14.

Other proteins or polypeptides may be attached to either the amino or carboxyl terminus of protein of the type illustrated in FIG. 2A. As an example, leader sequence 16 is shown extending from the amino terminal end of  $V_H$  domain 10.

FIG. 2B depicts another type of reagent 200 including a single polypeptide chain 100 and a pendant protein 18. Attached to the carboxyl end of the polypeptide chain 100 (which includes the FR and CDR sequences constituting an immunoglobulin binding site) is a pendant protein 18 consisting of, for example, a toxin or toxic fragment thereof, binding protein, enzyme or active enzyme fragment, or site of attachment for an imaging agent (e.g., to chelate a radioactive ion such as Indium-111).

FIG. 2C illustrates single chain polypeptide 300 including second single chain polypeptide 110 of the invention having the same or different specificity and connected via peptide linker 22 to the first single polypeptide chain 100.

FIG. 2D illustrates single chain polypeptide 400 which includes single polypeptide chains 110 and 100 linked together by linker 22, and pendant protein 18 attached to the carboxyl end of chain 110.



- 23 -

FIG. 2E illustrates single polypeptide chain 500 which includes chain 400 of Fig. 2D and pendant protein 20 (EGF) attached to the amino terminus of chain 400.

As is evident from Figs. 2A-E, single chain  
5 proteins of the invention may resemble beads on a string by including multiple biosynthetic binding sites, each binding site having unique specificity, or repeated sites of the same specificity to increase the avidity of the protein. As is evidenced from the  
10 foregoing, the invention provides a large family of reagents comprising proteins, at least a portion of which defines a binding site patterned after the variable region or regions of immunoglobulins to c-erbB-2 or related antigens.

15 The single chain polypeptides of the invention are designed at the DNA level. The synthetic DNAs are then expressed in a suitable host system, and the expressed proteins are collected and renatured if necessary.

20 The ability to design the single polypeptide chains of the invention depends on the ability to identify monoclonal antibodies of interest, and then to determine the sequence of the amino acids in the variable region of these antibodies, or the DNA  
25 sequence encoding them. Hybridoma technology enables production of cell lines secreting antibody to essentially any desired substance that elicits an immune response. For example, U.S. Patent No. 4,753,894 describes some monoclonal antibodies of  
30 interest which recognize c-erbB-2 related antigens on breast cancer cells, and explains how such antibodies were obtained. One monoclonal antibody that is particularly useful for this purpose is 520C9 (Bjorn et al. (1985) Cancer Res. 45:124-1221; U.S. Patent

- 24 -

No. 4,753,894). This antibody specifically recognizes the c-erbB-2 antigen expressed on the surface of various tumor cell lines, and exhibits very little binding to normal tissues. Alternative sources of sFv sequences with the desired specificity can take advantage of phage antibody and combinatorial library methodology. Such sequences would be based on cDNA from mice which were preimmunized with tumor cell membranes or c-erbB-2 or c-erbB-2-related antigenic fragments or peptides. (See, e.g., Clackson et al, Nature 352 624-628 (1991))

The process of designing DNA that encodes the single polypeptide chain of interest can be accomplished as follows. RNA encoding the light and heavy chains of the desired immunoglobulin can be obtained from the cytoplasm of the hybridoma producing the immunoglobulin. The mRNA can be used to prepare the cDNA for subsequent isolation of  $V_H$  and  $V_L$  genes by PCR methodology known in the art (Sambrook et al., eds., Molecular Cloning, 1989, Cold Spring Harbor Laboratories Press, NY). The N-terminal amino acid sequence of H and L chain may be independently determined by automated Edman sequencing; if necessary, further stretches of the CDRs and flanking FRs can be determined by amino acid sequencing of the H and L chain V region fragments. Such sequence analysis is now conducted routinely. This knowledge permits one to design synthetic primers for isolation of  $V_H$  and  $V_L$  genes from hybridoma cells that make monoclonal antibodies known to bind the c-erbB-2 or related antigen. These V genes will encode the Fv region that binds c-erbB-2 in the parent antibody.

Still another approach involves the design and construction of synthetic V genes that will encode an Fv binding site specific for c-erbB-2 or related

- 25 -

receptors. For example, with the help of a computer program such as, for example, Compugene, and known variable region DNA sequences, one may design and directly synthesize native or near-native FR sequences from a first antibody molecule, and CDR sequences from a second antibody molecule. The  $V_H$  and  $V_L$  sequences described above are linked together directly via an amino acid chain or linker connecting the C-terminus of one chain with the N-terminus of the other.

These genes, once synthesized, may be cloned with or without additional DNA sequences coding for, e.g., a leader peptide which facilitates secretion or intracellular stability of a fusion polypeptide, or a leader or trailing sequence coding for a second polypeptide. The genes then can be expressed directly in an appropriate host cell.

By directly sequencing an antibody to a c-erbB-2 or related antigen, or obtaining the sequence from the literature, in view of this disclosure, one skilled in the art can produce a single chain Fv comprising any desired CDR and FR. For example, using the DNA sequence for the 520C9 monoclonal antibody set forth in the Sequence Listing as SEQ ID NO:3, a single chain polypeptide can be produced having a binding affinity for a c-erbB-2 related antigen. Expressed sequences may be tested for binding and empirically refined by exchanging selected amino acids in relatively conserved regions, based on observation of trends in amino acid sequence data and/or computer modeling techniques. Significant flexibility in  $V_H$  and  $V_L$  design is possible because alterations in amino acid sequences may be made at the DNA level.

Accordingly, the construction of DNAs encoding the single-chain Fv and sFv fusion proteins of the

- 26 -

invention can be done using known techniques involving the use of various restriction enzymes which make sequence-specific cuts in DNA to produce blunt ends or cohesive ends, DNA ligases, techniques enabling enzymatic addition of sticky ends to blunt-ended DNA, construction of synthetic DNAs by assembly of short or medium length oligonucleotides, cDNA synthesis techniques, and synthetic probes for isolating immunoglobulin genes. Various promoter sequences and other regulatory RNA sequences used in achieving expression, and various type of host cells are also known and available. Conventional transfection techniques, and equally conventional techniques for cloning and subcloning DNA are useful in the practice of this invention and known to those skilled in the art. Various types of vectors may be used such as plasmids and viruses including animal viruses and bacteriophages. The vectors may exploit various marker genes which impart to a successfully transfected cell a detectable phenotypic property that can be used to identify which of a family of clones has successfully incorporated the recombinant DNA of the vector.

Of course, the processes for manipulating, amplifying, and recombining DNA which encode amino acid sequences of interest are generally well known in the art, and therefore, not described in detail herein. Methods of identifying the isolated V genes encoding antibody Fv regions of interest are well understood, and described in the patent and other literature. In general, the methods involve selecting genetic material coding for amino acid sequences which define the CDRs and FRs of interest upon reverse transcription, according to the genetic code.

- 27 -

One method of obtaining DNA encoding the single-chain Fv disclosed herein is by assembly of synthetic oligonucleotides produced in a conventional, automated, polynucleotide synthesizer followed by ligation with  
5 appropriate ligases. For example, overlapping, complementary DNA fragments comprising 15 bases may be synthesized semi-manually using phosphoramidite chemistry, with end segments left unphosphorylated to prevent polymerization during ligation. One end of the  
10 synthetic DNA is left with a "sticky end" corresponding to the site of action of a particular restriction endonuclease, and the other end is left with an end corresponding to the site of action of another restriction endonuclease. Alternatively, this approach  
15 can be fully automated. The DNA encoding the single chain polypeptides may be created by synthesizing longer single strand fragments (e.g., 50-100 nucleotides long) in, for example, a Biosearch oligonucleotide synthesizer, and then ligating the  
20 fragments.

Additional nucleotide sequences encoding, for example, constant region amino acids or a bioactive molecule may also be linked to the gene sequences to produce a bifunctional protein.

25 For example, the synthetic genes and DNA fragments designed as described above may be produced by assembly of chemically synthesized oligonucleotides. 15-100mer oligonucleotides may be synthesized on a Biosearch DNA Model 8600 Synthesizer, and purified by  
30 polyacrylamide gel electrophoresis (PAGE) in Tris-Borate-EDTA buffer (TBE). The DNA is then electroeluted from the gel. Overlapping oligomers may be phosphorylated by T4 polynucleotide kinase and ligated into larger blocks which may also be purified  
35 by PAGE.

- 28 -

The blocks or the pairs of longer oligonucleotides may be cloned in E. coli using a suitable cloning vector, e.g., pUC. Initially, this vector may be altered by single-strand mutagenesis to  
5 eliminate residual six base altered sites. For example,  $V_H$  may be synthesized and cloned into pUC as five primary blocks spanning the following restriction sites: (1) EcoRI to first NarI site; (2) first NarI to XbaI; (3) XbaI to SalI; (4) SalI to NcoI; and (5) NcoI  
10 to BamHI. These cloned fragments may then be isolated and assembled in several three-fragment ligations and cloning steps into the pUC8 plasmid. Desired ligations, selected by PAGE, are then transformed into, for example, E. coli strain JM83, and plated onto LB  
15 Ampicillin + Xgal plates according to standard procedures. The gene sequence may be confirmed by supercoil sequencing after cloning, or after subcloning into M13 via the dideoxy method of Sanger (Molecular Cloning, 1989, Sambrook et al., eds, 2d ed., Vol. 2,  
20 Cold Spring Harbor Laboratory Press, NY).

The engineered genes can be expressed in appropriate prokaryotic hosts such as various strains of E. coli, and in eucaryotic hosts such as Chinese hamster ovary cells (CHO), mouse myeloma, hybridoma,  
25 transfectoma, and human myeloma cells.

If the gene is to be expressed in E. coli, it may first be cloned into an expression vector. This is accomplished by positioning the engineered gene downstream from a promoter sequence such as Trp or Tac,  
30 and a gene coding for a leader polypeptide such as fragment B (FB) of staphylococcal protein A. The resulting expressed fusion protein accumulates in refractile bodies in the cytoplasm of the cells, and may be harvested after disruption of the cells by

- 29 -

French press or sonication. The refractile bodies are solubilized, and the expressed fusion proteins are cleaved and refolded by the methods already established for many other recombinant proteins (Huston et al, 5 1988, supra) or, for direct expression methods, there is no leader and the inclusion bodies may be refolded without cleavage (Huston et al, 1991, Methods in Enzymology, vol 203, pp 46-88).

For example, subsequent proteolytic cleavage of 10 the isolated sFv from their leader sequence fusions can be performed to yield free sFvs, which can be renatured to obtain an intact biosynthetic, hybrid antibody binding site. The cleavage site preferably is immediately adjacent the sFv polypeptide and includes 15 one amino acid or a sequence of amino acids exclusive of any one amino acid or amino acid sequence found in the amino acid structure of the single polypeptide chain.

The cleavage site preferably is designed for 20 specific cleavage by a selected agent. Endopeptidases are preferred, although non-enzymatic (chemical) cleavage agents may be used. Many useful cleavage agents, for instance, cyanogen bromide, dilute acid, trypsin, Staphylococcus aureus V-8 protease, post- 25 proline cleaving enzyme, blood coagulation Factor Xa, enterokinase, and renin, recognize and preferentially or exclusively cleave at particular cleavage sites. One currently preferred peptide sequence cleavage agent is V-8 protease. The currently preferred cleavage site 30 is at a Glu residue. Other useful enzymes recognize multiple residues as a cleavage site, e.g., factor Xa (Ile-Glu-Gly-Arg) or enterokinase (Asp-Asp-Asp-Asp-Lys). Dilute acid preferentially leaves the peptide bond between Asp-Pro residues, and CNBr in acid cleaves 35 after Met, unless it is followed by Tyr.

- 30 -

If the engineered gene is to be expressed in eucaryotic hybridoma cells, the conventional expression system for immunoglobulins, it is first inserted into an expression vector containing, for example, the immunoglobulin promoter, a secretion signal, immunoglobulin enhancers, and various introns. This plasmid may also contain sequences encoding another polypeptide such as all or part of a constant region, enabling an entire part of a heavy or light chain to be expressed, or at least part of a toxin, enzyme, cytokine, or hormone. The gene is transfected into myeloma cells via established electroporation or protoplast fusion methods. Cells so transfected may then express  $V_H$ -linker- $V_L$  or  $V_L$ -linker- $V_H$  single-chain Fv polypeptides, each of which may be attached in the various ways discussed above to a protein domain having another function (e.g., cytotoxicity).

For construction of a single contiguous chain of amino acids specifying multiple binding sites, restriction sites at the boundaries of DNA encoding a single binding site (i.e.,  $V_H$ -linker- $V_L$ ) are utilized or created, if not already present. DNAs encoding single binding sites are ligated and cloned into shuttle plasmids, from which they may be further assembled and cloned into the expression plasmid. The order of domains will be varied and spacers between the domains provide flexibility needed for independent folding of the domains. The optimal architecture with respect to expression levels, refolding and functional activity will be determined empirically. To create bivalent sFv's, for example, the stop codon in the gene encoding the first binding site is changed to an open reading frame, and several glycine plus serine codons including a restriction site such as BamHI (encoding



- 31 -

Gly-Ser) or XhoI (encoding Gly-Ser-Ser) are put in place. The second sFv gene is modified similarly at its 5' end, receiving the same restriction site in the same reading frame. The genes are combined at this site to produce the bivalent sFv gene.

Linkers connecting the C-terminus of one domain to the N-terminus of the next generally comprise hydrophilic amino acids which assume an unstructured configuration in physiological solutions and preferably are free of residues having large side groups which might interfere with proper folding of the  $V_H$ ,  $V_L$ , or pendant chains. One useful linker has the amino acid sequence  $[(Gly)_4Ser]_3$  (see SEQ ID NOS:5 and 6, residue numbers 121-135). One currently preferred linker has the amino acid sequence comprising 2 or 3 repeats of  $[(Ser)_4Gly]$ , such as  $[(Ser)_4Gly]_2$  and  $[(Ser)_4Gly]_3$  (see SEQ ID NOS:3 and 4).

The invention is illustrated further by the following non-limiting Examples.

20

#### EXAMPLES

##### 1. Antibodies to c-erbB-2 Related Antigens

Monoclonal antibodies against breast cancer have been developed using human breast cancer cells or membrane extracts of the cells for immunizing mice, as described in Frankel et al. (1985) J. Biol. Resp. Modif. 4:273-286, hereby incorporated by reference. Hybridomas have been made and selected for production of antibodies using a panel of normal and breast cancer cells. A panel of eight normal tissue membranes, a fibroblast cell line, and frozen sections of breast cancer tissues were used in the screening. Candidates that passed the first screening were further tested on 16 normal tissue sections, 5 normal blood cell types,

- 32 -

11 nonbreast neoplasm sections, 21 breast cancer sections, and 14 breast cancer cell lines. From this selection, 127 antibodies were selected. Irrelevant antibodies and nonbreast cancer cell lines were used in control experiments.

Useful monoclonal antibodies were found to include 520C9, 454C11 (A.T.C.C. Nos. HB8696 and HB8484, respectively) and 741F8. Antibodies identified as selective for breast cancer in this screen reacted against five different antigens. The sizes of the antigens that the antibodies recognize: 200 kD; a series of proteins that are probably degradation products with Mr's of 200 kD, 93kD, 60 kD, and 37 kD; 180 kD (transferrin receptor); 42 kD; and 55 kD, respectively. Of the antibodies directed against the five classes of antigens, the most specific are the ones directed against the 200 kD antigen, 520C9 being a representative antibody for that antigen class. 520C9 reacts with fewer breast cancer tissues (about 20-70% depending on the assay conditions) and it reacts with the fewest normal tissues of any of the antibodies. 520C9 reacts with kidney tubules (as do many monoclonal antibodies), but not pancreas, esophagus, lung, colon, stomach, brain, tonsil, liver, heart, ovary, skin, bone, uterus, bladder, or normal breast among some of the tissues tested.

2. Preparation of cDNA Library Encoding 520C9 Antibody.

Polyadenylated RNA was isolated from approximately  $1 \times 10^8$  (520C9 hybridoma) cells using the "FAST TRACK" mRNA isolation kit from Invitrogen (San Diego, CA). The presence of immunoglobulin heavy chain RNA was confirmed by Northern analysis (Molecular Cloning, 1989, Sambrook et al., eds., 2d ed., Cold

- 33 -

Spring Harbor Laboratory Press, NY) using a recombinant probe containing the various J regions of heavy chain genomic DNA. Using 6  $\mu$ g RNA for each, cDNA was prepared using the Invitrogen cDNA synthesis system with either random and oligo dT primers. Following synthesis, the cDNA was size-selected by isolating 0.5-3.0 Kilobase (Kb) fragments following agarose gel electrophoresis. After optimizing the cDNA to vector ratio, these fragments were then ligated to the pcDNA II Invitrogen cloning vector.

3. Isolation of V<sub>H</sub> and V<sub>L</sub> Domains

After transformation of the bacteria with plasmid library DNA, colony hybridization was performed using antibody constant (C) region and joining (J) region probes for either light or heavy chain genes. See Orlandi, R., et al., 1989, Proc. Nat. Aca. Sci. 86:3833. The antibody constant region probe can be obtained from any of light or heavy chain nucleotide sequences from an immunoglobulin gene using known procedures. Several potential positive clones were identified for both heavy and light chain genes and, after purification by a second round of screening, these were sequenced. One clone (M207) contained the sequence of non-functional Kappa chain which has a tyrosine substituted for a conserved cysteine, and also terminates prematurely due to a 4 base deletion which causes a frame-shift mutation in the variable-J region junction. A second light chain clone (M230) contained virtually the entire 520C9 light chain gene except for the last 18 amino acids of the constant region and approximately half of the signal sequence. The 520C9 heavy chain variable region was present on a clone of approximately 1,100 base pairs (F320) which ended near the end of the CH2 domain.

- 34 -

4. Mutagenesis of V<sub>H</sub> AND V<sub>L</sub>

In order to construct the sFv, both the heavy and light chain variable regions were mutagenized to insert appropriate restriction sites (Kunkel, T.A., 1985, Proc. Nat. Acad. Sci. USA 82:1373). The heavy chain clone (F320) was mutagenized to insert a BamHI site at the 5' end of V<sub>H</sub> (F321). The light chain was also mutagenized simultaneously by inserting an EcoRV site at the 5' end and a PstI site with a translation stop codon at the 3' end of the variable region (M231).

5. Sequencing

cDNA clones encoding light and heavy chain were sequenced using external standard pUC primers and several specific internal primers which were prepared on the basis of the sequences obtained for the heavy chain. The nucleotide sequences were analyzed in a Genbank homology search (program Nucscan of DNA-star) to eliminate endogenous immunoglobulin genes. Translation into amino acids was checked with amino acid sequences in the NIH atlas edited by E. Kabat.

Amino acid sequences derived from 520C9 immunoglobulin confirmed the identity of these V<sub>H</sub> and V<sub>L</sub> cDNA clones. The heavy chain clone pF320 started 6 nucleotides upstream of the first ATG codon and extended into the CH2-encoding region, but it lacked the last nine amino acid codons of the CH2 constant domain and all of the CH3 coding region, as well as the 3' untranslated region and the poly A tail. Another short heavy chain clone containing only the CH2 and CH3 coding regions, and the poly A tail was initially assumed to represent the missing part of the 520C9 heavy chain. However, overlap between both sequences was not identical. The 520C9 clone (pF320) encodes the CH1 and CH2 domains of murine IgG1, whereas the short clone pF315 encodes the CH2 and CH3 of IgG2b.

- 35 -

## 6. Gene Design

A nucleic acid sequence encoding a composite 520C9 sFv region containing a single-chain Fv binding site which recognizes c-erbB-2 related tumor antigens was designed with the aid of Compugene software. The gene contains nucleic acid sequences encoding the V<sub>H</sub> and V<sub>L</sub> regions of the 520C9 antibody described above linked together with a double-stranded synthetic oligonucleotide coding for a peptide with the amino acid sequence set forth in the Sequence Listing as amino acid residue numbers 116 through 133 in SEQ ID NOS:3 and 4. This linker oligonucleotide contains helper cloning sites EcoRI and BamHI, and was designed to contain the assembly sites SacI and EcoRV near its 5' and 3' ends, respectively. These sites enable match-up and ligation to the 3' and 5' ends of 520C9 V<sub>H</sub> and V<sub>L</sub>, respectively, which also contain these sites (V<sub>H</sub>-linker-V<sub>L</sub>). However, the order of linkage to the oligonucleotide may be reversed (V<sub>L</sub>-linker-V<sub>H</sub>) in this or any sFv of the invention. Other restriction sites were designed into the gene to provide alternative assembly sites. A sequence encoding the FB fragment of protein A was used as a leader.

The invention also embodies a humanized single-chain Fv, i.e., containing human framework sequences and CDR sequences which specify c-erbB-2 binding, e.g., like the CDRs of the 520C9 antibody. The humanized Fv is thus capable of binding c-erbB-2 while eliciting little or no immune response when administered to a patient. A nucleic acid sequence encoding a humanized sFv may be designed and constructed as follows. Two strategies for sFv design are especially useful. A homology search in the GenBank database for the most related human framework (FR) regions may be performed

- 36 -

and FR regions of the sFv may be mutagenized according to sequences identified in the search to reproduce the corresponding human sequence; or information from computer modeling based on x-ray structures of model Fab fragments may be used (Amit et al., 1986, Science 233:747-753; Colman et al., 1987, Nature 326:358-363; Sheriff et al., 1987, Proc. Nat. Aca. Sci., 84:8075-8079; and Satow et al., 1986, J. Mol. Biol. 190:593-604, all of which are hereby incorporated by reference). In a preferred case, the most homologous human  $V_H$  and  $V_L$  sequences may be selected from a collection of PCR-cloned human V regions. The FRs are made synthetically and fused to CDRs to make successively more complete V regions by PCR-based ligation, until the full humanized  $V_L$  and  $V_H$  are completed. For example, a humanized sFv that is a hybrid of the murine 520C9 antibody CDRs and the human myeloma protein NEW FRs can be designed such that each variable region has the murine binding site within a human framework (FR1-CDR1-FR2-CDR2-FR3-CDR3-FR4). The Fab NEW crystal structure (Saul et al., 1978, J. Biol. Chem. 253:585-597) also may be used to predict the location of FRs in the variable regions. Once these regions are predicted, the amino acid sequence or the corresponding nucleotide sequence of the regions may be determined, and the sequences may be synthesized and cloned into shuttle plasmids, from which they may be further assembled and cloned into an expression plasmid; alternatively, the FR sequences of the 520C9 sFv may be mutagenized directly and the changes verified by supercoil sequencing with internal primers (Chen et al., 1985, DNA 4:165-170).

- 37 -

7. Preparation of and Purification 520C9 sFv

## A. Inclusion Body Solubilization.

The 520C9 sFv plasmid, based on a T<sub>7</sub> promoter and vector, was made by direct expression in E. coli of the fused gene sequence set forth in the Sequence Listing as SEQ. ID NO:3. Inclusion bodies (15.8 g) from a 2.0 liter fermentation were washed with 25 mM Tris, 10 mM EDTA, pH 8.0 (TE), plus 1 M guanidine hydrochloride (GuHCl). The inclusion bodies were solubilized in TE, 6 M GuHCl, 10 mM dithiothreitol (DTT), pH 9.0, and yielded 3825 A<sub>280</sub> units of material. This material was ethanol precipitated, washed with TE, 3M urea, then resuspended in TE, 8M urea, 10 mM DTT, pH 8.0. This precipitation step prepared the protein for ion exchange purification of the denatured sFv.

## B. Ion Exchange Chromatography

The solubilized inclusion bodies were subjected to ion exchange chromatography in an effort to remove contaminating nucleic acids and E. coli proteins before renaturation of the sFv. The solubilized inclusion bodies in 8M urea were diluted with TE to a final urea concentration of 6M, then passed through 100 ml of DEAE-Sepharose Fast Flow in a radial flow column. The sFv was recovered in the unbound fraction (69% of the starting sample).

The pH of this sFv solution (A<sub>280</sub> = 5.7; 290 ml) was adjusted to 5.5 with 1 M acetic acid to prepare it for application to an S-Sepharose Fast Flow column. When the pH went below 6.0, however, precipitate formed in the sample. The sample was clarified; 60% of the sample was in the pellet and 40% in the supernatant. The supernatant was passed through 100 ml S-Sepharose Fast Flow and the sFv recovered in the unbound fraction. The pellet was resolubilized in TE, 6 M

- 38 -

GuHCl, 10 mM DTT, pH 9.0, and was also found to contain primarily sFv in a pool of 45 ml volume with an absorbance at 280 nm of 20 absorbance units. This reduced sFv pool was carried through the remaining  
5 steps of the purification.

#### C. Renaturation of sFv

Renaturation of the sFv was accomplished using a disulfide-restricted refolding approach, in which the disulfides were oxidized while the sFv was fully  
10 denatured, followed by removal of the denaturant and refolding. Oxidation of the sFv samples was carried out in TE, 6 M GuHCl, 1 mM oxidized glutathione (GSSG), 0.1 mM reduced glutathione (GSH), pH 9.0. The sFv was diluted into the oxidation buffer to a final protein  
15  $A_{280} = 0.075$  with a volume of 4000 ml and incubated overnight at room temperature. After overnight oxidation this solution was dialyzed against 10 mM sodium phosphate, 1 mM EDTA, 150 mM NaCl, 500 mM urea, pH 8.0 (PENU) [4 x (20 liters X 24 hrs)]. Low levels  
20 of activity were detected in the refolded sample.

#### D. Membrane Fractionation and Concentration of Active sFv

In order to remove aggregated misfolded material before any concentration step, the dialyzed refolded  
25 520C9 sFv (5050 ml) was filtered through a 100K MWCO membrane (100,000 mol. wt. cut-off) (4 x 60 cm<sup>2</sup>) using a Minitan ultrafiltration device (Millipore). This step required a considerable length of time (9 hours), primarily due to formation of precipitate in the  
30 retentate and membrane fouling as the protein concentration in the retentate increased. 95% of the protein in the refolded sample was retained by the 100K membranes, with 79% in the form of insoluble material. The 100K retentate had very low activity and was  
35 discarded.



- 39 -

The 100K filtrate contained most of the soluble sFv activity for binding c-erbB-2, and it was next concentrated using 10K MWCO membranes (10,000 mol. wt. cut-off) (4 x 60 cm<sup>2</sup>) in the Minitan, to a volume of 100 ml (50X). This material was further concentrated using a YM10 10K MWCO membrane in a 50 ml Amicon stirred cell to a final volume of 5.2 ml (1000X). Only a slight amount of precipitate formed during the two 10K concentration steps. The specific activity of this concentrated material was significantly increased relative to the initial dialyzed refolding.

#### E. Size Exclusion Chromatography of Concentrated sFv

When refolded sFv was fractionated by size exclusion chromatography, all 520C9 sFv activity was determined to elute at the position of folded monomer. In order to enrich for active monomers, the 1000X concentrated sFv sample was fractionated on a Sephacryl S-200 HR column (2.5 x 40 cm) in PBSA (2.7 mM KCl, 1.1 mM KH<sub>2</sub>PO<sub>4</sub>, 138 mM NaCl, 8.1 mM Na<sub>2</sub>HPO<sub>4</sub> · 7H<sub>2</sub>O, 0.02% NaN<sub>3</sub>) + 0.5 M urea. The elution profile of the column and SDS-PAGE analysis of the fractions showed two sFv monomer peaks. The two sFv monomer peak fractions were pooled (10 ml total) and displayed c-erbB-2 binding activity in competition assays.

#### F. Affinity Purification of 520C9 sFv

The extracellular domain of (ECD) c-erbB-2 was expressed in baculovirus-infected insect cells. This protein (ECD c-erbB-2) was immobilized on an agarose affinity matrix. The sFv monomer peak was dialyzed against PBSA to remove the urea and then applied to a 0.7 x 4.5 cm ECD c-erbB-2-agarose affinity column in PBSA. The column was washed to baseline A<sub>280</sub>, then eluted with PBSA + 3 M LiCl, pH = 6.1. The peak

- 40 -

fractions were pooled (4 ml) and dialyzed against PBSA to remove the LiCl. 72  $\mu$ g of purified sFv was obtained from 750  $\mu$ g of S-200 monomer fractions. Activity measurements on the column fractions were determined by a competitive assay. Briefly, sFv affinity purification fractions and HRP-conjugated 520C9 Fab fragments were allowed to compete for binding to SK-BR-3 membranes. Successful binding of the sFv preparation prevented the HRP-52069 Fab fragment from binding to the membranes, thus also reducing or preventing utilization of the HRP substrate, and no color development (see below for details of competition assay). The results showed that virtually all of the sFv activity was bound by the column and was recovered in the eluted peak (Figure 4). As expected, the specific activity of the eluted peak was increased relative to the column sample, and appeared to be essentially the same as the parent Fab control, within the experimental error of these measurements.

20 9. Yield After Purification.

Table I shows the yield of various 520C9 preparations during the purification process. Protein concentration ( $\mu$ g/ml) was determined by the BioRad protein assay. Under "Total Yield", 300 AU denatured sFv stock represents 3.15 g inclusion bodies from 0.4 liters fermentation. The oxidation buffer was 25 mM Tris, 10 mM EDTA, 6 M GdnHCl, 1 mM GSSG, 0.1 mM GSH, pH 9.0. Oxidation was performed at room temperature overnight. Oxidized sample was dialyzed against 10 mM sodium phosphate, 1 mM EDTA, 150 mM NaCl, 500 mM urea, pH 8.0. All subsequent steps were carried out in this buffer, except for affinity chromatography, which was carried out in PBSA.

- 41 -

Table I

	<u>Sample</u>	<u>Volume</u>	<u>Protein Concentration</u>	<u>Total Yield</u>	<u>% Yield</u>
5	1. Refolding III (oxidation)	4000 ml	0.075 A <sub>280</sub>	300 AU	-
10	2. Dialyzed Refolding III	5050 ml	38 µg/ml	191.9 mg	100
	3. Minitan 100K Filtrate	5000 ml	2 µg/ml	10.0 mg	5.4
15	4. Minitan 10K Retentate	100 ml	45 µg/ml	4.5 mg	2.3
	6. YM10 10K Retentate	5.2 ml	600 µg/ml	3.1 mg	1.6
20	7. S-200 sFv Monomer Peak	10.0 ml	58 µg/ml	0.58 mg	0.3
25	8. Affinity Purified sFv	5.5 ml	13 µg/ml	0.07 mg	0.04

- 42 -

10. Immunotoxin Construction

The ricin A-520C9 single chain fused immunotoxin (SEQ. ID NO:7) encoding gene was constructed by isolating the gene coding for ricin A on a HindIII to BamHI fragment from pPL229 (Cetus Corporation, Emeryville, CA) and using it upstream of the 520C9 sFv in pH777, as shown in FIG. 3. This fusion contains the 122 amino acid natural linker present between the A and B domains of ricin. However, in the original pRAP229 expression vector the codon for amino acid 268 of ricin was converted to a TAA translation stop codon so that the expression of the resulting gene produces only ricin A. Therefore, in order to remove the translation stop codon, site-directed mutagenesis was performed to remove the TAA and restore the natural serine codon. This then allows translation to continue through the entire immunotoxin gene.

In order to insert the immunotoxin back into the pPL229 and pRAP229 expression vectors, the PstI site at the end of the immunotoxin gene had to be converted to a sequence that was compatible with the BamHI site in vector. A synthetic oligonucleotide adaptor containing a BclI site nested between PstI ends was inserted. BclI and BamHI ends are compatible and can be combined into a hybrid BclI/BamHI site. Since BclI nuclease is sensitive to dam methylation, the construction first was transformed into a dam(-) E. coli strain, Gm48, in order to digest the plasmid DNA with BclI (and HindIII), then insert the entire immunotoxin gene on a HindIII/BclI fragment back into both Hind III/BamHI-digested expression vectors.

When native 520C9 IgG1 is conjugated with native ricin A chain or recombinant ricin A chain, the resulting immunotoxin is able to inhibit protein

- 43 -

synthesis by 50% at a concentration of about  $0.4 \times 10^{-9}$  M against SK-Br-3 cells. In addition to reacting with SK-Br-3 breast cancer cells, native 520C9 IgG1 immunotoxin also inhibits an ovarian cancer cell line, OVCAR-3, with a  $ID_{50}$  of  $2.0 \times 10^{-9}$  M.

In the ricin A-sFv fusion protein described above, ricin acts as leader for expression, i.e., is fused to the amino terminus of sFv. Following direct expression, soluble protein was shown to react with antibodies against native 520C9 Fab and also to exhibit ricin A chain enzymatic activity.

In another design, the ricin A chain is fused to the carboxy terminus of sFv. The 520C9 sFv may be secreted via the PelB signal sequence with ricin A chain attached to the C-terminus of sFv. For this construct, sequences encoding the PelB-signal sequence, sFv, and ricin are joined in a bluescript plasmid via a HindIII site directly following sFv (in our expression plasmids) and the HindIII site preceding the ricin gene, in a three part assembly (RI-HindIII-BamHI). A new PstI site following the ricin gene is obtained via the Bluescript polylinker. Mutagenesis of this DNA removes the stop codon and the original PstI site at the end of sFv, and places several serine residues between the sFv and ricin genes. This new gene fusion, PelB signal sequence/sFv/ricin A, can be inserted into expression vectors as an EcoRI/PstI fragment.

In another design, the pseudomonas exotoxin fragment analogous to ricin A chain, PE40, is fused to the carboxy terminus of the anti-c-erbB-2 741F8 sFv (Seq ID NOS: 15 and 16). The resulting 741F8 sFv-PE40 is a single-chain Fv-toxin fusion protein, which was constructed with an 18 residue short FB leader which initially was left on the protein. E. coli expression

- 44 -

of this protein produced inclusion bodies that were refolded in a 3 M urea glutathione/redox buffer. The resulting sFv-PE40 was shown to specifically kill c-erbB-2 bearing cells in culture more fully and with apparently better cytotoxicity than the corresponding crosslinked immunotoxin. The sFv-toxin protein, as well as the 741F8 sFv, can be made in good yields by these procedures, and may be used as therapeutic and diagnostic agents for tumors bearing the c-erbB-2 or related antigens, such as breast and ovarian cancer.

#### 11. Assays

##### A. Competition ELISA

SK-Br-3 extract is prepared as a source of c-erbB-2 antigen as follows. SK-Br-3 breast cancer cells (Ring et al. 1989, Cancer Research 49:3070-3080), are grown to near confluence in Iscove's medium (Gibco BRL, Gaithersburg, Md.) plus 5% fetal bovine serum and 2 mM glutamine. The medium is aspirated, and the cells are rinsed with 10 ml fetal bovine serum (FBS) plus calcium and magnesium. The cells are scraped off with a rubber policeman into 10 ml FBS plus calcium and magnesium, and the flask is rinsed out with another 5 ml of this buffer. The cells are then centrifuged at 100 rpm. The supernate is aspirated off, and the cells are resuspended at  $10^7$  cells/ml in 10 mM NaCl, 0.5% NP40, pH 8 (TNN buffer), and are pipetted up and down to dissolve the pellet. The solution is then centrifuged at 1000 rpm to remove nuclei and other insoluble debris. The extract is filtered through 0.45 Millex HA and 0.2 Millex Gv filters. The TNN extract is stored as aliquots in Wheaton freezing vials at -70°C.

A fresh vial of SK-Br-3 TNN extract is thawed and diluted 200-fold into deionized water. Immediately thereafter, 40ug per well are added to a Dynatech PVC

- 45 -

96 well plak, which is allowed to sit overnight in a 37°C dry incubator. The plates are washed four times in phosphate buffered saline (PBS), 1% skim milk, 0.05% Tween 20.

- 5           The non-specific binding sites are blocked as follows. When the plate is dry, 100 ug per well PBS is added containing 1% skim milk, and the incubation allowed to proceed for one hour at room temperature.

- The single-chain Fv test samples and standard  
10 520C9 whole antibody dilutions are then added as follows. 520C9 antibody and test samples are diluted in dilution buffer (PBS + 1% skim milk) in serial two-fold steps, initially at 50ug/ml and making at least 10  
15 dilutions for 520C9 standards. A control containing only dilution buffer is included. The diluted samples and standards are added at 50ul per well and incubated for 30 minutes at room temperature.

- The 520C9-horseradish peroxidase (HRP) probe is added as follows. 520C9-HRP conjugate (Zymed Labs.,  
20 South San Francisco, California) is diluted to 14 ug/ml with 1% skim milk in dilution buffer. The optimum dilutions must be determined for each new batch of peroxidase conjugate without removing the previous steps. 20 ul per well of probe was added and incubated  
25 for one hour at room temperature. The plate is then washed four times in PBS. The peroxidase substrate is then added. The substrate solution should be made fresh for each use by diluting tetramethyl benzidine stock (TMB; 2mg/ml in 100% ethanol) 1:20 and 3%  
30 hydrogen peroxide stock 1:2200 in substrate buffer (10mM sodium acetate, 10mM Na, EDTA, pH 5.0). This is incubated for 30 minutes at room temperature. The wells are then quenched with 100 ul per well 0.8 M H<sub>2</sub>SO<sub>4</sub> and the absorbance at 150 nm read.

- 46 -

FIG. 4 compares the binding ability of the parent refolded but unpurified 520C9 monoclonal antibody, 520C9 Fab fragments, and the 520C9 sFv single-chain binding site after binding and elution from an affinity column (eluted) or the unbound flow through fraction (passed). In Fig. 4, the fully purified 520C9 sFv exhibits an affinity for c-erbB-2 that is indistinguishable from the parent monoclonal antibody, within the error of measuring protein concentration.

10       B. In vivo testing

Immunotoxins that are strong inhibitors of protein synthesis against breast cancer cells grown in culture may be tested for their in vivo efficacy. The in vivo assay is typically done in a nude mouse model using xenografts of human MX-1 breast cancer cells. Mice are injected with either PBS (control) or different concentrations of sFv-toxin immunotoxin, and a concentration-dependent inhibition of tumor growth will be observed. It is expected that higher doses of immunotoxin will produce a better effect.

The invention may be embodied in other specific forms without departing from the spirit and scope thereof. The present embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the meaning and range of equivalence of the claims are intended to be embraced therein.

30



- 47 -

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Huston, James S.  
Oppermann, Hermann  
Houston, L. L.  
Ring, David B.
- (ii) TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer Marker
- (iii) NUMBER OF SEQUENCES: 16
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Edmund R. Pitcher, Testa, Hurwitz, & Thibeault
  - (B) STREET: Exchange Place, 53 State Street
  - (C) CITY: Boston
  - (D) STATE: Massachusetts
  - (E) COUNTRY: USA
  - (F) ZIP: 02109
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE:
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Pitcher, Edmund R.
  - (B) REGISTRATION NUMBER: 27,829
  - (C) REFERENCE/DOCKET NUMBER: 2054/22
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: (617) 248-7000
  - (B) TELEFAX: (617) 248-7100

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 4299 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA

- 48 -

## (ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..4299

(D) OTHER INFORMATION: /note= "product = "c-erb-b-2""

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG GAG CTG GCG GCC TTG TGC CGC TGG GGG CTC CTC CTC GCC CTC TTG	48
Met Glu Leu Ala Ala Leu Cys Arg Trp Gly Leu Leu Leu Ala Leu Leu	
1 5 10 15	
CCC CCC GGA GCC GCG AGC ACC CAA GTG TGC ACC GGC ACA GAC ATG AAG	96
Pro Pro Gly Ala Ala Ser Thr Gln Val Cys Thr Gly Thr Asp Met Lys	
20 25 30	
CTG CGG CTC CCT GCC AGT CCC GAG ACC CAC CTG GAC ATG CTC CGC CAC	144
Leu Arg Leu Pro Ala Ser Pro Glu Thr His Leu Asp Met Leu Arg His	
35 40 45	
CTC TAC CAG GGC TGC CAG GTG GTG CAG GGA AAC CTG GAA CTC ACC TAC	192
Leu Tyr Gln Gly Cys Gln Val Val Gln Gly Asn Leu Glu Leu Thr Tyr	
50 55 60	
CTG CCC ACC AAT GCC AGC CTG TCC TTC CTG CAG GAT ATC CAG GAG GTG	240
Leu Pro Thr Asn Ala Ser Leu Ser Phe Leu Gln Asp Ile Gln Glu Val	
65 70 75 80	
CAG GGC TAC GTG CTC ATC GCT CAC AAC CAA GTG AGG CAG GTC CCA CTG	288
Gln Gly Tyr Val Leu Ile Ala His Asn Gln Val Arg Gln Val Pro Leu	
85 90 95	
CAG AGG CTG CGG ATT GTG CGA GGC ACC CAG CTC TTT GAG GAC AAC TAT	336
Gln Arg Leu Arg Ile Val Arg Gly Thr Gln Leu Phe Glu Asp Asn Tyr	
100 105 110	
GCC CTG GCC GTG CTA GAC AAT GGA GAC CCG CTG AAC AAT ACC ACC CCT	384
Ala Leu Ala Val Leu Asp Asn Gly Asp Pro Leu Asn Asn Thr Thr Pro	
115 120 125	
GTC ACA GGG GCC TCC CCA GGA GGC CTG CGG GAG CTG CAG CTT CGA AGC	432
Val Thr Gly Ala Ser Pro Gly Gly Leu Arg Glu Leu Gln Leu Arg Ser	
130 135 140	
CTC ACA GAG ATC TTG AAA GGA GGG GTC TTG ATC CAG CGG AAC CCC CAG	480
Leu Thr Glu Ile Leu Lys Gly Gly Val Leu Ile Gln Arg Asn Pro Gln	
145 150 155 160	
CTC TGC TAC CAG GAC ACG ATT TTG TGG AAG GAC ATC TTC CAC AAG AAC	528
Leu Cys Tyr Gln Asp Thr Ile Leu Trp Lys Asp Ile Phe His Lys Asn	
165 170 175	
AAC CAG CTG GCT CTC ACA CTG ATA GAC ACC AAC CGC TCT CGG GCC TGC	576
Asn Gln Leu Ala Leu Thr Leu Ile Asp Thr Asn Arg Ser Arg Ala Cys	
180 185 190	

- 49 -

CAC	CCC	TGT	TCT	CCG	ATG	TGT	AAG	GGC	TCC	CGC	TGC	TGG	GGA	GAG	AGT	624
His	Pro	Cys	Ser	Pro	Met	Cys	Lys	Gly	Ser	Arg	Cys	Trp	Gly	Glu	Ser	
		195					200					205				
TCT	GAG	GAT	TGT	CAG	AGC	CTG	ACG	CGC	ACT	GTC	TGT	GCC	GGT	GGC	TGT	672
Ser	Glu	Asp	Cys	Gln	Ser	Leu	Thr	Arg	Thr	Val	Cys	Ala	Gly	Gly	Cys	
	210					215					220					
GCC	CGC	TGC	AAG	GGG	CCA	CTG	CCC	ACT	GAC	TGC	TGC	CAT	GAG	CAG	TGT	720
Ala	Arg	Cys	Lys	Gly	Pro	Leu	Pro	Thr	Asp	Cys	Cys	His	Glu	Gln	Cys	
225					230					235					240	
GCT	GCC	GGC	TGC	ACG	GGC	CCC	AAG	CAC	TCT	GAC	TGC	CTG	GCC	TGC	CTC	768
Ala	Ala	Gly	Cys	Thr	Gly	Pro	Lys	His	Ser	Asp	Cys	Leu	Ala	Cys	Leu	
				245					250					255		
CAC	TTC	AAC	CAC	AGT	GGC	ATC	TGT	GAG	CTG	CAC	TGC	CCA	GCC	CTG	GTC	816
His	Phe	Asn	His	Ser	Gly	Ile	Cys	Glu	Leu	His	Cys	Pro	Ala	Leu	Val	
			260					265					270			
ACC	TAC	AAC	ACA	GAC	ACG	TTT	GAG	TCC	ATG	CCC	AAT	CCC	GAG	GGC	CGG	864
Thr	Tyr	Asn	Thr	Asp	Thr	Phe	Glu	Ser	Met	Pro	Asn	Pro	Glu	Gly	Arg	
		275					280					285				
TAT	ACA	TTC	GGC	GCC	AGC	TGT	GTG	ACT	GCC	TGT	CCC	TAC	AAC	TAC	CTT	912
Tyr	Thr	Phe	Gly	Ala	Ser	Cys	Val	Thr	Ala	Cys	Pro	Tyr	Asn	Tyr	Leu	
	290					295					300					
TCT	ACG	GAC	GTG	GGA	TCC	TGC	ACC	CTC	GTC	TGC	CCC	CTG	CAC	AAC	CAA	960
Ser	Thr	Asp	Val	Gly	Ser	Cys	Thr	Leu	Val	Cys	Pro	Leu	His	Asn	Gln	
305					310					315					320	
GAG	GTG	ACA	GCA	GAG	GAT	GGA	ACA	CAG	CGG	TGT	GAG	AAG	TGC	AGC	AAG	1008
Glu	Val	Thr	Ala	Glu	Asp	Gly	Thr	Gln	Arg	Cys	Glu	Lys	Cys	Ser	Lys	
				325					330					335		
CCC	TGT	GCC	CGA	GTG	TGC	TAT	GGT	CTG	GGC	ATG	GAG	CAC	TTG	CGA	GAG	1056
Pro	Cys	Ala	Arg	Val	Cys	Tyr	Gly	Leu	Gly	Met	Glu	His	Leu	Arg	Glu	
			340					345					350			
GTG	AGG	GCA	GTT	ACC	AGT	GCC	AAT	ATC	CAG	GAG	TTT	GCT	GGC	TGC	AAG	1104
Val	Arg	Ala	Val	Thr	Ser	Ala	Asn	Ile	Gln	Glu	Phe	Ala	Gly	Cys	Lys	
		355					360					365				
AAG	ATC	TTT	GGG	AGC	CTG	GCA	TTT	CTG	CCG	GAG	AGC	TTT	GAT	GGG	GAC	1152
Lys	Ile	Phe	Gly	Ser	Leu	Ala	Phe	Leu	Pro	Glu	Ser	Phe	Asp	Gly	Asp	
	370					375					380					
CCA	GCC	TCC	AAC	ACT	GCC	CCG	CTC	CAG	CCA	GAG	CAG	CTC	CAA	GTG	TTT	1200
Pro	Ala	Ser	Asn	Thr	Ala	Pro	Leu	Gln	Pro	Glu	Gln	Leu	Gln	Val	Phe	
385					390					395					400	
GAG	ACT	CTG	GAA	GAG	ATC	ACA	GGT	TAC	CTA	TAC	ATC	TCA	GCA	TGG	CCG	1248
Glu	Thr	Leu	Glu	Glu	Ile	Thr	Gly	Tyr	Leu	Tyr	Ile	Ser	Ala	Trp	Pro	
				405					410					415		

- 50 -

GAC	AGC	CTG	CCT	GAC	CTC	AGC	GTC	TTC	CAG	AAC	CTG	CAA	GTA	ATC	CGG	1296
Asp	Ser	Leu	Pro	Asp	Leu	Ser	Val	Phe	Gln	Asn	Leu	Gln	Val	Ile	Arg	
			420					425					430			
GGA	CGA	ATT	CTG	CAC	AAT	GGC	GCC	TAC	TCG	CTG	ACC	CTG	CAA	GGG	CTG	1344
Gly	Arg	Ile	Leu	His	Asn	Gly	Ala	Tyr	Ser	Leu	Thr	Leu	Gln	Gly	Leu	
		435					440					445				
GGC	ATC	AGC	TGG	CTG	GGG	CTG	CGC	TCA	CTG	AGG	GAA	CTG	GGC	AGT	GGA	1392
Gly	Ile	Ser	Trp	Leu	Gly	Leu	Arg	Ser	Leu	Arg	Glu	Leu	Gly	Ser	Gly	
	450					455					460					
CTG	GCC	CTC	ATC	CAC	CAT	AAC	ACC	CAC	CTC	TGC	TTC	GTG	CAC	ACG	GTG	1440
Leu	Ala	Leu	Ile	His	His	Asn	Thr	His	Leu	Cys	Phe	Val	His	Thr	Val	
465					470				475						480	
CCC	TGG	GAC	CAG	CTC	TTT	CGG	AAC	CCG	CAC	CAA	GCT	CTG	CTC	CAC	ACT	1488
Pro	Trp	Asp	Gln	Leu	Phe	Arg	Asn	Pro	His	Gln	Ala	Leu	Leu	His	Thr	
			485					490						495		
GCC	AAC	CGG	CCA	GAG	GAC	GAG	TGT	GTG	GGC	GAG	GGC	CTG	GCC	TGC	CAC	1536
Ala	Asn	Arg	Pro	Glu	Asp	Glu	Cys	Val	Gly	Glu	Gly	Leu	Ala	Cys	His	
			500					505					510			
CAG	CTG	TGC	GCC	CGA	GGG	CAC	TGC	TGG	GGT	CCA	GGG	CCC	ACC	CAG	TGT	1584
Gln	Leu	Cys	Ala	Arg	Gly	His	Cys	Trp	Gly	Pro	Gly	Pro	Thr	Gln	Cys	
		515					520					525				
GTC	AAC	TGC	AGC	CAG	TTC	CTT	CGG	GGC	CAG	GAG	TGC	GTG	GAG	GAA	TGC	1632
Val	Asn	Cys	Ser	Gln	Phe	Leu	Arg	Gly	Gln	Glu	Cys	Val	Glu	Glu	Cys	
	530					535					540					
CGA	GTA	CTG	CAG	GGG	CTC	CCC	AGG	GAG	TAT	GTG	AAT	GCC	AGG	CAC	TGT	1680
Arg	Val	Leu	Gln	Gly	Leu	Pro	Arg	Glu	Tyr	Val	Asn	Ala	Arg	His	Cys	
545					550					555					560	
TTG	CCG	TGC	CAC	CCT	GAG	TGT	CAG	CCC	CAG	AAT	GGC	TCA	GTG	ACC	TGT	1728
Leu	Pro	Cys	His	Pro	Glu	Cys	Gln	Pro	Gln	Asn	Gly	Ser	Val	Thr	Cys	
				565				570						575		
TTT	GGA	CCG	GAG	GCT	GAC	CAG	TGT	GTG	GCC	TGT	GCC	CAC	TAT	AAG	GAC	1776
Phe	Gly	Pro	Glu	Ala	Asp	Gln	Cys	Val	Ala	Cys	Ala	His	Tyr	Lys	Asp	
			580					585					590			
CCT	CCC	TTC	TGC	GTG	GCC	CGC	TGC	CCC	AGC	GGT	GTG	AAA	CCT	GAC	CTC	1824
Pro	Pro	Phe	Cys	Val	Ala	Arg	Cys	Pro	Ser	Gly	Val	Lys	Pro	Asp	Leu	
		595					600					605				
TCC	TAC	ATG	CCC	ATC	TGG	AAG	TTT	CCA	GAT	GAG	GAG	GGC	GCA	TGC	CAG	1872
Ser	Tyr	Met	Pro	Ile	Trp	Lys	Phe	Pro	Asp	Glu	Glu	Gly	Ala	Cys	Gln	
	610					615					620					
CCT	TGC	CCC	ATC	AAC	TGC	ACC	CAC	TCC	TGT	GTG	GAC	CTG	GAT	GAC	AAG	1920
Pro	Cys	Pro	Ile	Asn	Cys	Thr	His	Ser	Cys	Val	Asp	Leu	Asp	Asp	Lys	
625					630					635					640	

- 51 -

GGC TGC CCC GCC GAG CAG AGA GCC AGC CCT CTG ACG TCC ATC ATC TCT	1968
Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu Thr Ser Ile Ile Ser	
645 650 655	
GCG GTG GTT GGC ATT CTG CTG GTC GTG GTC TTG GGG GTG GTC TTT GGG	2016
Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Val Phe Gly	
660 665 670	
ATC CTC ATC AAG CGA CGG CAG CAG AAG ATC CGG AAG TAC ACG ATG CGG	2064
Ile Leu Ile Lys Arg Arg Gln Gln Lys Ile Arg Lys Tyr Thr Met Arg	
675 680 685	
AGA CTG CTG CAG GAA ACG GAG CTG GTG GAG CCG CTG ACA CCT AGC GGA	2112
Arg Leu Leu Gln Glu Thr Glu Leu Val Glu Pro Leu Thr Pro Ser Gly	
690 695 700	
GCG ATG CCC AAC CAG GCG CAG ATG CGG ATC CTG AAA GAG ACG GAG CTG	2160
Ala Met Pro Asn Gln Ala Gln Met Arg Ile Leu Lys Glu Thr Glu Leu	
705 710 715 720	
AGG AAG GTG AAG GTG CTT GGA TCT GGC GCT TTT GGC ACA GTC TAC AAG	2208
Arg Lys Val Lys Val Leu Gly Ser Gly Ala Phe Gly Thr Val Tyr Lys	
725 730 735	
GGC ATC TGG ATC CCT GAT GGG GAG AAT GTG AAA ATT CCA GTG GCC ATC	2256
Gly Ile Trp Ile Pro Asp Gly Glu Asn Val Lys Ile Pro Val Ala Ile	
740 745 750	
AAA GTG TTG AGG GAA AAC ACA TCC CCC AAA GCC AAC AAA GAA ATC TTA	2304
Lys Val Leu Arg Glu Asn Thr Ser Pro Lys Ala Asn Lys Glu Ile Leu	
755 760 765	
GAC GAA GCA TAC GTG ATG GCT GGT GTG GGC TCC CCA TAT GTC TCC CGC	2352
Asp Glu Ala Tyr Val Met Ala Gly Val Gly Ser Pro Tyr Val Ser Arg	
770 775 780	
CTT CTG GGC ATC TGC CTG ACA TCC ACG GTG CAG CTG GTG ACA CAG CTT	2400
Leu Leu Gly Ile Cys Leu Thr Ser Thr Val Gln Leu Val Thr Gln Leu	
785 790 795 800	
ATG CCC TAT GGC TGC CTC TTA GAC CAT GTC CGG GAA AAC CGC GGA CGC	2448
Met Pro Tyr Gly Cys Leu Leu Asp His Val Arg Glu Asn Arg Gly Arg	
805 810 815	
CTG GGC TCC CAG GAC CTG CTG AAC TGG TGT ATG CAG ATT GCC AAG GGG	2496
Leu Gly Ser Gln Asp Leu Leu Asn Trp Cys Met Gln Ile Ala Lys Gly	
820 825 830	
ATG AGC TAC CTG GAG GAT GTG CGG CTC GTA CAC AGG GAC TTG GCC GCT	2544
Met Ser Tyr Leu Glu Asp Val Arg Leu Val His Arg Asp Leu Ala Ala	
835 840 845	
CGG AAC GTG CTG GTC AAG AGT CCC AAC CAT GTC AAA ATT ACA GAC TTC	2592
Arg Asn Val Leu Val Lys Ser Pro Asn His Val Lys Ile Thr Asp Phe	
850 855 860	

- 52 -

GGG CTG GCT CGG CTG CTG GAC ATT GAC GAG ACA GAG TAC CAT GCA GAT	2640
Gly Leu Ala Arg Leu Leu Asp Ile Asp Glu Thr Glu Tyr His Ala Asp	
865 870 875 880	
GGG GGC AAG GTG CCC ATC AAG TGG ATG GCG CTG GAG TCC ATT CTC CGC	2688
Gly Gly Lys Val Pro Ile Lys Trp Met Ala Leu Glu Ser Ile Leu Arg	
885 890 895	
CGG CGG TTC ACC CAC CAG AGT GAT GTG TGG AGT TAT GGT GTG ACT GTG	2736
Arg Arg Phe Thr His Gln Ser Asp Val Trp Ser Tyr Gly Val Thr Val	
900 905 910	
TGG GAG CTG ATG ACT TTT GGG GCC AAA CCT TAC GAT GGG ATC CCA GCC	2784
Trp Glu Leu Met Thr Phe Gly Ala Lys Pro Tyr Asp Gly Ile Pro Ala	
915 920 925	
CGG GAG ATC CCT GAC CTG CTG GAA AAG GGG GAG CGG CTG CCC CAG CCC	2832
Arg Glu Ile Pro Asp Leu Leu Glu Lys Gly Glu Arg Leu Pro Gln Pro	
930 935 940	
CCC ATC TGC ACC ATT GAT GTC TAC ATG ATC ATG GTC AAA TGT TGG ATG	2880
Pro Ile Cys Thr Ile Asp Val Tyr Met Ile Met Val Lys Cys Trp Met	
945 950 955 960	
ATT GAC TCT GAA TGT CGG CCA AGA TTC CGG GAG TTG GTG TCT GAA TTC	2928
Ile Asp Ser Glu Cys Arg Pro Arg Phe Arg Glu Leu Val Ser Glu Phe	
965 970 975	
TCC CGC ATG GCC AGG GAC CCC CAG CGC TTT GTG GTC ATC CAG AAT GAG	2976
Ser Arg Met Ala Arg Asp Pro Gln Arg Phe Val Val Ile Gln Asn Glu	
980 985 990	
GAC TTG GGC CCA GCC AGT CCC TTG GAC AGC ACC TTC TAC CGC TCA CTG	3024
Asp Leu Gly Pro Ala Ser Pro Leu Asp Ser Thr Phe Tyr Arg Ser Leu	
995 1000 1005	
CTG GAG GAC GAT GAC ATG GGG GAC CTG GTG GAT GCT GAG GAG TAT CTG	3072
Leu Glu Asp Asp Asp Met Gly Asp Leu Val Asp Ala Glu Glu Tyr Leu	
1010 1015 1020	
GTA CCC CAG CAG GGC TTC TTC TGT CCA GAC CCT GCC CCG GGC GCT GGG	3120
Val Pro Gln Gln Gly Phe Phe Cys Pro Asp Pro Ala Pro Gly Ala Gly	
1025 1030 1035 1040	
GGC ATG GTC CAC CAC AGG CAC CGC AGC TCA TCT ACC AGG AGT GGC GGT	3168
Gly Met Val His His Arg His Arg Ser Ser Ser Thr Arg Ser Gly Gly	
1045 1050 1055	
GGG GAC CTG ACA CTA GGG CTG GAG CGC TCT GAA GAG GAG GCC CCC AGG	3216
Gly Asp Leu Thr Leu Gly Leu Glu Pro Ser Glu Glu Glu Ala Pro Arg	
1060 1065 1070	
TCT CCA CTG GCA CCC TCC GAA GGG GCT GGC TCC GAT GTA TTT GAT GGT	3264
Ser Pro Leu Ala Pro Ser Glu Gly Ala Gly Ser Asp Val Phe Asp Gly	
1075 1080 1085	

- 53 -

GAC CTG GGA ATG GGG GCA GCC AAG GGG CTG CAA AGC CTC CCC ACA CAT Asp Leu Gly Met Gly Ala Ala Lys Gly Leu Gln Ser Leu Pro Thr His 1090 1095 1100	3312
GAC CCC AGC CCT CTA CAG CGG TAC AGT GAG GAC CCC ACA GTA CCC CTG Asp Pro Ser Pro Leu Gln Arg Tyr Ser Glu Asp Pro Thr Val Pro Leu 1105 1110 1115 1120	3360
CCC TCT GAG ACT GAT GGC TAC GTT GCC CCC CTG ACC TGC AGC CCC CAG Pro Ser Glu Thr Asp Gly Tyr Val Ala Pro Leu Thr Cys Ser Pro Gln 1125 1130 1135	3408
CCT GAA TAT GTG AAC CAG CCA GAT GTT CGG CCC CAG CCC CCT TCG CCC Pro Glu Tyr Val Asn Gln Pro Asp Val Arg Pro Gln Pro Pro Ser Pro 1140 1145 1150	3456
CGA GAG GGC CCT CTG CCT GCT GCC CGA CCT GCT GGT GCC ACT CTG GAA Arg Glu Gly Pro Leu Pro Ala Ala Arg Pro Ala Gly Ala Thr Leu Glu 1155 1160 1165	3504
AGG CCC AAG ACT CTC TCC CCA GGG AAG AAT GGG GTC GTC AAA GAC GTT Arg Pro Lys Thr Leu Ser Pro Gly Lys Asn Gly Val Val Lys Asp Val 1170 1175 1180	3552
TTT GCC TTT GGG GGT GCC GTG GAG AAC CCC GAG TAC TTG ACA CCC CAG Phe Ala Phe Gly Gly Ala Val Glu Asn Pro Glu Tyr Leu Thr Pro Gln 1185 1190 1195 1200	3600
GGA GGA GCT GCC CCT CAG CCC CAC CCT CCT CCT GCC TTC AGC CCA GCC Gly Gly Ala Ala Pro Gln Pro His Pro Pro Pro Ala Phe Ser Pro Ala 1205 1210 1215	3648
TTC GAC AAC CTC TAT TAC TGG GAC CAG GAC CCA CCA GAG CGG GGG GCT Phe Asp Asn Leu Tyr Tyr Trp Asp Gln Asp Pro Pro Glu Arg Gly Ala 1220 1225 1230	3696
CCA CCC AGC ACC TTC AAA GGG ACA CCT ACG GCA GAG AAC CCA GAG TAC Pro Pro Ser Thr Phe Lys Gly Thr Pro Thr Ala Glu Asn Pro Glu Tyr 1235 1240 1245	3744
CTG GGT CTG GAC GTG CCA GTG TGA ACC AGA AGG CCA AGT CCG CAG AAG Leu Gly Leu Asp Val Pro Val * Thr Arg Arg Pro Ser Pro Gln Lys 1250 1255 1260	3792
CCC TGA TGT GTC CTC AGG GAG CAG GGA AGG CCT GAC TTC TGC TGG CAT Pro * Cys Val Leu Arg Glu Gln Gly Arg Pro Asp Phe Cys Trp His 1265 1270 1275 1280	3840
CAA GAG GTG GGA GGG CCC TCC GAC CAC TTC CAG GGG AAC CTG CCA TGC Gln Glu Val Gly Gly Pro Ser Asp His Phe Gln Gly Asn Leu Pro Cys 1285 1290 1295	3888

- 54 -

CAG GAA CCT GTC CTA AGG AAC CTT CCT TCC TGC TTG AGT TCC CAG ATG Gln Glu Pro Val Leu Arg Asn Leu Pro Ser Cys Leu Ser Ser Gln Met 1300 1305 1310	3936
GCT GGA AGG GGT CCA GCC TCG TTG GAA GAG GAA CAG CAC TGG GGA GTC Ala Gly Arg Gly Pro Ala Ser Leu Glu Glu Glu Gln His Trp Gly Val 1315 1320 1325	3984
TTT GTG GAT TCT GAG GCC CTG CCC AAT GAG ACT CTA GGG TCC AGT GGA Phe Val Asp Ser Glu Ala Leu Pro Asn Glu Thr Leu Gly Ser Ser Gly 1330 1335 1340	4032
TGC CAC AGC CCA GCT TGG CCC TTT CCT TCC AGA TCC TGG GTA CTG AAA Cys His Ser Pro Ala Trp Pro Phe Pro Ser Arg Ser Trp Val Leu Lys 1345 1350 1355 1360	4080
GCC TTA GGG AAG CTG GCC TGA GAG GGG AAG CGG CCC TAA GGG AGT GTC Ala Leu Gly Lys Leu Ala * Glu Gly Lys Arg Pro * Gly Ser Val 1365 1370 1375	4128
TAA GAA CAA AAG CGA CCC ATT CAG AGA CTG TCC CTG AAA CCT AGT ACT * Glu Gln Lys Arg Pro Ile Gln Arg Leu Ser Leu Lys Pro Ser Thr 1380 1385 1390	4176
GCC CCC CAT GAG GAA GGA ACA GCA ATG GTG TCA GTA TCC AGG CTT TGT Ala Pro His Glu Glu Gly Thr Ala Met Val Ser Val Ser Arg Leu Cys 1395 1400 1405	4224
ACA GAG TGC TTT TCT GTT TAG TTT TTA CTT TTT TTG TTT TGT TTT TTT Thr Glu Cys Phe Ser Val * Phe Leu Leu Phe Leu Phe Cys Phe Phe 1410 1415 1420	4272
AAA GAT GAA ATA AAG ACC CAG GGG GAG Lys Asp Glu Ile Lys Thr Gln Gly Glu 1425 1430	4299

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1433 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Glu	Leu	Ala	Ala	Leu	Cys	Arg	Trp	Gly	Leu	Leu	Leu	Ala	Leu	Leu
1				5					10					15	
Pro	Pro	Gly	Ala	Ala	Ser	Thr	Gln	Val	Cys	Thr	Gly	Thr	Asp	Met	Lys
			20					25					30		



Leu	Arg	Leu	Pro	Ala	Ser	Pro	Glu	Thr	His	Leu	Asp	Met	Leu	Arg	His
35						40				45					
Leu	Tyr	Gln	Gly	Cys	Gln	Val	Val	Gln	Gly	Asn	Leu	Glu	Leu	Thr	Tyr
50						55				60					
Leu	Pro	Thr	Asn	Ala	Ser	Leu	Ser	Phe	Leu	Gln	Asp	Ile	Gln	Glu	Val
65				70						75				80	
Gln	Gly	Tyr	Val	Leu	Ile	Ala	His	Asn	Gln	Val	Arg	Gln	Val	Pro	Leu
				85				90						95	
Gln	Arg	Leu	Arg	Ile	Val	Arg	Gly	Thr	Gln	Leu	Phe	Glu	Asp	Asn	Tyr
		100						105				110			
Ala	Leu	Ala	Val	Leu	Asp	Asn	Gly	Asp	Pro	Leu	Asn	Asn	Thr	Thr	Pro
115						120						125			
Val	Thr	Gly	Ala	Ser	Pro	Gly	Gly	Leu	Arg	Glu	Leu	Gln	Leu	Arg	Ser
130						135				140					
Leu	Thr	Glu	Ile	Leu	Lys	Gly	Gly	Val	Leu	Ile	Gln	Arg	Asn	Pro	Gln
145				150						155				160	
Leu	Cys	Tyr	Gln	Asp	Thr	Ile	Leu	Trp	Lys	Asp	Ile	Phe	His	Lys	Asn
				165				170						175	
Asn	Gln	Leu	Ala	Leu	Thr	Leu	Ile	Asp	Thr	Asn	Arg	Ser	Arg	Ala	Cys
		180						185				190			
His	Pro	Cys	Ser	Pro	Met	Cys	Lys	Gly	Ser	Arg	Cys	Trp	Gly	Glu	Ser
		195				200						205			
Ser	Glu	Asp	Cys	Gln	Ser	Leu	Thr	Arg	Thr	Val	Cys	Ala	Gly	Gly	Cys
210						215				220					
Ala	Arg	Cys	Lys	Gly	Pro	Leu	Pro	Thr	Asp	Cys	Cys	His	Glu	Gln	Cys
225				230						235				240	
Ala	Ala	Gly	Cys	Thr	Gly	Pro	Lys	His	Ser	Asp	Cys	Leu	Ala	Cys	Leu
				245				250						255	
His	Phe	Asn	His	Ser	Gly	Ile	Cys	Glu	Leu	His	Cys	Pro	Ala	Leu	Val
		260						265				270			
Thr	Tyr	Asn	Thr	Asp	Thr	Phe	Glu	Ser	Met	Pro	Asn	Pro	Glu	Gly	Arg
		275				280						285			
Tyr	Thr	Phe	Gly	Ala	Ser	Cys	Val	Thr	Ala	Cys	Pro	Tyr	Asn	Tyr	Leu
290						295				300					
Ser	Thr	Asp	Val	Gly	Ser	Cys	Thr	Leu	Val	Cys	Pro	Leu	His	Asn	Gln
305				310						315				320	

- 56 -

Glu Val Thr Ala Glu Asp Gly Thr Gln Arg Cys Glu Lys Cys Ser Lys  
 325 330 335  
 Pro Cys Ala Arg Val Cys Tyr Gly Leu Gly Met Glu His Leu Arg Glu  
 340 345 350  
 Val Arg Ala Val Thr Ser Ala Asn Ile Gln Glu Phe Ala Gly Cys Lys  
 355 360 365  
 Lys Ile Phe Gly Ser Leu Ala Phe Leu Pro Glu Ser Phe Asp Gly Asp  
 370 375 380  
 Pro Ala Ser Asn Thr Ala Pro Leu Gln Pro Glu Gln Leu Gln Val Phe  
 385 390 395 400  
 Glu Thr Leu Glu Glu Ile Thr Gly Tyr Leu Tyr Ile Ser Ala Trp Pro  
 405 410 415  
 Asp Ser Leu Pro Asp Leu Ser Val Phe Gln Asn Leu Gln Val Ile Arg  
 420 425 430  
 Gly Arg Ile Leu His Asn Gly Ala Tyr Ser Leu Thr Leu Gln Gly Leu  
 435 440 445  
 Gly Ile Ser Trp Leu Gly Leu Arg Ser Leu Arg Glu Leu Gly Ser Gly  
 450 455 460  
 Leu Ala Leu Ile His His Asn Thr His Leu Cys Phe Val His Thr Val  
 465 470 475 480  
 Pro Trp Asp Gln Leu Phe Arg Asn Pro His Gln Ala Leu Leu His Thr  
 485 490 495  
 Ala Asn Arg Pro Glu Asp Glu Cys Val Gly Glu Gly Leu Ala Cys His  
 500 505 510  
 Gln Leu Cys Ala Arg Gly His Cys Trp Gly Pro Gly Pro Thr Gln Cys  
 515 520 525  
 Val Asn Cys Ser Gln Phe Leu Arg Gly Gln Glu Cys Val Glu Glu Cys  
 530 535 540  
 Arg Val Leu Gln Gly Leu Pro Arg Glu Tyr Val Asn Ala Arg His Cys  
 545 550 555 560  
 Leu Pro Cys His Pro Glu Cys Gln Pro Gln Asn Gly Ser Val Thr Cys  
 565 570 575  
 Phe Gly Pro Glu Ala Asp Gln Cys Val Ala Cys Ala His Tyr Lys Asp  
 580 585 590  
 Pro Pro Phe Cys Val Ala Arg Cys Pro Ser Gly Val Lys Pro Asp Leu  
 595 600 605

- 57 -

Ser Tyr Met Pro Ile Trp Lys Phe Pro Asp Glu Glu Gly Ala Cys Gln  
 610 615 620  
 Pro Cys Pro Ile Asn Cys Thr His Ser Cys Val Asp Leu Asp Asp Lys  
 625 630 635 640  
 Gly Cys Pro Ala Glu Gln Arg Ala Ser Pro Leu Thr Ser Ile Ile Ser  
 645 650 655  
 Ala Val Val Gly Ile Leu Leu Val Val Val Leu Gly Val Val Phe Gly  
 660 665 670  
 Ile Leu Ile Lys Arg Arg Gln Gln Lys Ile Arg Lys Tyr Thr Met Arg  
 675 680 685  
 Arg Leu Leu Gln Glu Thr Glu Leu Val Glu Pro Leu Thr Pro Ser Gly  
 690 695 700  
 Ala Met Pro Asn Gln Ala Gln Met Arg Ile Leu Lys Glu Thr Glu Leu  
 705 710 715 720  
 Arg Lys Val Lys Val Leu Gly Ser Gly Ala Phe Gly Thr Val Tyr Lys  
 725 730 735  
 Gly Ile Trp Ile Pro Asp Gly Glu Asn Val Lys Ile Pro Val Ala Ile  
 740 745 750  
 Lys Val Leu Arg Glu Asn Thr Ser Pro Lys Ala Asn Lys Glu Ile Leu  
 755 760 765  
 Asp Glu Ala Tyr Val Met Ala Gly Val Gly Ser Pro Tyr Val Ser Arg  
 770 775 780  
 Leu Leu Gly Ile Cys Leu Thr Ser Thr Val Gln Leu Val Thr Gln Leu  
 785 790 795 800  
 Met Pro Tyr Gly Cys Leu Leu Asp His Val Arg Glu Asn Arg Gly Arg  
 805 810 815  
 Leu Gly Ser Gln Asp Leu Leu Asn Trp Cys Met Gln Ile Ala Lys Gly  
 820 825 830  
 Met Ser Tyr Leu Glu Asp Val Arg Leu Val His Arg Asp Leu Ala Ala  
 835 840 845  
 Arg Asn Val Leu Val Lys Ser Pro Asn His Val Lys Ile Thr Asp Phe  
 850 855 860  
 Gly Leu Ala Arg Leu Leu Asp Ile Asp Glu Thr Glu Tyr His Ala Asp  
 865 870 875 880  
 Gly Gly Lys Val Pro Ile Lys Trp Met Ala Leu Glu Ser Ile Leu Arg  
 885 890 895

- 58 -

Arg Arg Phe Thr His Gln Ser Asp Val Trp Ser Tyr Gly Val Thr Val  
 900 905 910

Trp Glu Leu Met Thr Phe Gly Ala Lys Pro Tyr Asp Gly Ile Pro Ala  
 915 920 925

Arg Glu Ile Pro Asp Leu Leu Glu Lys Gly Glu Arg Leu Pro Gln Pro  
 930 935 940

Pro Ile Cys Thr Ile Asp Val Tyr Met Ile Met Val Lys Cys Trp Met  
 945 950 955 960

Ile Asp Ser Glu Cys Arg Pro Arg Phe Arg Glu Leu Val Ser Glu Phe  
 965 970 975

Ser Arg Met Ala Arg Asp Pro Gln Arg Phe Val Val Ile Gln Asn Glu  
 980 985 990

Asp Leu Gly Pro Ala Ser Pro Leu Asp Ser Thr Phe Tyr Arg Ser Leu  
 995 1000 1005

Leu Glu Asp Asp Asp Met Gly Asp Leu Val Asp Ala Glu Glu Tyr Leu  
 1010 1015 1020

Val Pro Gln Gln Gly Phe Phe Cys Pro Asp Pro Ala Pro Gly Ala Gly  
 1025 1030 1035 1040

Gly Met Val His His Arg His Arg Ser Ser Ser Thr Arg Ser Gly Gly  
 1045 1050 1055

Gly Asp Leu Thr Leu Gly Leu Glu Pro Ser Glu Glu Glu Ala Pro Arg  
 1060 1065 1070

Ser Pro Leu Ala Pro Ser Glu Gly Ala Gly Ser Asp Val Phe Asp Gly  
 1075 1080 1085

Asp Leu Gly Met Gly Ala Ala Lys Gly Leu Gln Ser Leu Pro Thr His  
 1090 1095 1100

Asp Pro Ser Pro Leu Gln Arg Tyr Ser Glu Asp Pro Thr Val Pro Leu  
 1105 1110 1115 1120

Pro Ser Glu Thr Asp Gly Tyr Val Ala Pro Leu Thr Cys Ser Pro Gln  
 1125 1130 1135

Pro Glu Tyr Val Asn Gln Pro Asp Val Arg Pro Gln Pro Pro Ser Pro  
 1140 1145 1150

Arg Glu Gly Pro Leu Pro Ala Ala Arg Pro Ala Gly Ala Thr Leu Glu  
 1155 1160 1165

Arg Pro Lys Thr Leu Ser Pro Gly Lys Asn Gly Val Val Lys Asp Val  
 1170 1175 1180

- 59 -

Phe Ala Phe Gly Gly Ala Val Glu Asn Pro Glu Tyr Leu Thr Pro Gln  
 1185 1190 1195 1200  
 Gly Gly Ala Ala Pro Gln Pro His Pro Pro Pro Ala Phe Ser Pro Ala  
 1205 1210 1215  
 Phe Asp Asn Leu Tyr Tyr Trp Asp Gln Asp Pro Pro Glu Arg Gly Ala  
 1220 1225 1230  
 Pro Pro Ser Thr Phe Lys Gly Thr Pro Thr Ala Glu Asn Pro Glu Tyr  
 1235 1240 1245  
 Leu Gly Leu Asp Val Pro Val \* Thr Arg Arg Pro Ser Pro Gln Lys  
 1250 1255 1260  
 Pro \* Cys Val Leu Arg Glu Gln Gly Arg Pro Asp Phe Cys Trp His  
 1265 1270 1275 1280  
 Gln Glu Val Gly Gly Pro Ser Asp His Phe Gln Gly Asn Leu Pro Cys  
 1285 1290 1295  
 Gln Glu Pro Val Leu Arg Asn Leu Pro Ser Cys Leu Ser Ser Gln Met  
 1300 1305 1310  
 Ala Gly Arg Gly Pro Ala Ser Leu Glu Glu Glu Gln His Trp Gly Val  
 1315 1320 1325  
 Phe Val Asp Ser Glu Ala Leu Pro Asn Glu Thr Leu Gly Ser Ser Gly  
 1330 1335 1340  
 Cys His Ser Pro Ala Trp Pro Phe Pro Ser Arg Ser Trp Val Leu Lys  
 1345 1350 1355 1360  
 Ala Leu Gly Lys Leu Ala \* Glu Gly Lys Arg Pro \* Gly Ser Val  
 1365 1370 1375  
 \* Glu Gln Lys Arg Pro Ile Gln Arg Leu Ser Leu Lys Pro Ser Thr  
 1380 1385 1390  
 Ala Pro His Glu Glu Gly Thr Ala Met Val Ser Val Ser Arg Leu Cys  
 1395 1400 1405  
 Thr Glu Cys Phe Ser Val \* Phe Leu Leu Phe Leu Phe Cys Phe Phe  
 1410 1415 1420  
 Lys Asp Glu Ile Lys Thr Gln Gly Glu  
 1425 1430

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 739 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

- 60 -

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..739

(D) OTHER INFORMATION: /note= "product = "520C9sFv/ amino acid info: 520C9sFv protein"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAG ATC CAA TTG GTG CAG TCT GGA CCT GAG CTG AAG AAG CCT GGA GAG	48
Glu Ile Gln Leu Val Gln Ser Gly Pro Glu Leu Lys Lys Pro Gly Glu	
1 5 10 15	
ACA GTC AAG ATC TCC TGC AAG GCT TCT GGA TAT ACC TTC GCA AAC TAT	96
Thr Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Ala Asn Tyr	
20 25 30	
GGA ATG AAC TGG ATG AAG CAG GCT CCA GGA AAG GGT TTA AAG TGG ATG	144
Gly Met Asn Trp Met Lys Gln Ala Pro Gly Lys Gly Leu Lys Trp Met	
35 40 45	
GGC TGG ATA AAC ACC TAC ACT GGA CAG TCA ACA TAT GCT GAT GAC TTC	192
Gly Trp Ile Asn Thr Tyr Thr Gly Gln Ser Thr Tyr Ala Asp Asp Phe	
50 55 60	
AAG GAA CGG TTT GCC TTC TCT TTG GAA ACC TCT GCC ACC ACT GCC CAT	240
Lys Glu Arg Phe Ala Phe Ser Leu Glu Thr Ser Ala Thr Thr Ala His	
65 70 75 80	
TTG CAG ATC AAC AAC CTC AGA AAT GAG GAC TCG GCC ACA TAT TTC TGT	288
Leu Gln Ile Asn Asn Leu Arg Asn Glu Asp Ser Ala Thr Tyr Phe Cys	
85 90 95	
GCA AGA CGA TTT GGG TTT GCT TAC TGG GGC CAA GGG ACT CTG GTC AGT	336
Ala Arg Arg Phe Gly Phe Ala Tyr Trp Gly Gln Gly Thr Leu Val Ser	
100 105 110	
GTC TCT GCA TCG ATA TCG AGC TCC TCC GGA TCT TCA TCT AGC GGT TCC	384
Val Ser Ala Ser Ile Ser Ser Ser Ser Gly Ser Ser Ser Ser Gly Ser	
115 120 125	
AGC TCG AGT GGA TCC GAT ATC CAG ATG ACC CAG TCT CCA TCC TCC TTA	432
Ser Ser Ser Gly Ser Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu	
130 135 140	
TCT GCC TCT CTG GGA GAA AGA GTC AGT CTC ACT TGT CGG GCA AGT CAG	480
Ser Ala Ser Leu Gly Glu Arg Val Ser Leu Thr Cys Arg Ala Ser Gln	
145 150 155 160	
GAC ATT GGT AAT AGC TTA ACC TGG CTT CAG CAG GAA CCA GAT GGA ACT	528
Asp Ile Gly Asn Ser Leu Thr Trp Leu Gln Gln Glu Pro Asp Gly Thr	
165 170 175	

- 61 -

ATT AAA CGC CTG ATC TAC GCC ACA TCC AGT TTA GAT TCT GGT GTC CCC	576
Ile Lys Arg Leu Ile Tyr Ala Thr Ser Ser Leu Asp Ser Gly Val Pro	
180 185 190	
AAA AGG TTC AGT GGC AGT CGG TCT GGG TCA GAT TAT TCT CTC ACC ATC	624
Lys Arg Phe Ser Gly Ser Arg Ser Gly Ser Asp Tyr Ser Leu Thr Ile	
195 200 205	
AGT AGC CTT GAG TCT GAA GAT TTT GTA GTC TAT TAC TGT CTA CAA TAT	672
Ser Ser Leu Glu Ser Glu Asp Phe Val Val Tyr Tyr Cys Leu Gln Tyr	
210 215 220	
GCT ATT TTT CCG TAC ACG TTC GGA GGG GGG ACC AAC CTG GAA ATA AAA	720
Ala Ile Phe Pro Tyr Thr Phe Gly Gly Gly Thr Asn Leu Glu Ile Lys	
225 230 235 240	
CGG GCT GAT TAA TCT GCA G	739
Arg Ala Asp * Ser Ala	
245	

## (2) INFORMATION FOR SEQ ID NO:4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 246 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Glu	Ile	Gln	Leu	Val	Gln	Ser	Gly	Pro	Glu	Leu	Lys	Lys	Pro	Gly	Glu
1				5					10					15	
Thr	Val	Lys	Ile	Ser	Cys	Lys	Ala	Ser	Gly	Tyr	Thr	Phe	Ala	Asn	Tyr
		20					25						30		
Gly	Met	Asn	Trp	Met	Lys	Gln	Ala	Pro	Gly	Lys	Gly	Leu	Lys	Trp	Met
		35					40					45			
Gly	Trp	Ile	Asn	Thr	Tyr	Thr	Gly	Gln	Ser	Thr	Tyr	Ala	Asp	Asp	Phe
	50					55					60				
Lys	Glu	Arg	Phe	Ala	Phe	Ser	Leu	Glu	Thr	Ser	Ala	Thr	Thr	Ala	His
	65				70					75					80
Leu	Gln	Ile	Asn	Asn	Leu	Arg	Asn	Glu	Asp	Ser	Ala	Thr	Tyr	Phe	Cys
				85				90						95	
Ala	Arg	Arg	Phe	Gly	Phe	Ala	Tyr	Trp	Gly	Gln	Gly	Thr	Leu	Val	Ser
			100				105						110		
Val	Ser	Ala	Ser	Ile	Ser	Ser	Ser	Gly	Ser	Ser	Ser	Ser	Gly	Ser	
		115					120					125			

- 62 -

Ser Ser Ser Gly Ser Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu  
 130 135 140

Ser Ala Ser Leu Gly Glu Arg Val Ser Leu Thr Cys Arg Ala Ser Gln  
 145 150 155 160

Asp Ile Gly Asn Ser Leu Thr Trp Leu Gln Gln Glu Pro Asp Gly Thr  
 165 170 175

Ile Lys Arg Leu Ile Tyr Ala Thr Ser Ser Leu Asp Ser Gly Val Pro  
 180 185 190

Lys Arg Phe Ser Gly Ser Arg Ser Gly Ser Asp Tyr Ser Leu Thr Ile  
 195 200 205

Ser Ser Leu Glu Ser Glu Asp Phe Val Val Tyr Tyr Cys Leu Gln Tyr  
 210 215 220

Ala Ile Phe Pro Tyr Thr Phe Gly Gly Gly Thr Asn Leu Glu Ile Lys  
 225 230 235 240

Arg Ala Asp \* Ser Ala  
 245

- (2) INFORMATION FOR SEQ ID NO:5: DELETED ACCORDING TO  
 PRELIMINARY AMENDMENT
- (2) INFORMATION FOR SEQ ID NO:6: DELETED ACCORDING TO  
 PRELIMINARY AMENDMENT
- (2) INFORMATION FOR SEQ IS NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 807 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS  
 (B) LOCATION: 1..807  
 (D) OTHER INFORMATION: /note= "product = "Ricin-A chain  
 gene/ amino acid info: Ricin-A chain protein"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ATG	ATA	TTC	CCC	AAA	CAA	TAC	CCA	ATT	ATA	AAC	TTT	ACC	ACA	GCG	GGT	48
Met	Ile	Phe	Pro	Lys	Gln	Tyr	Pro	Ile	Ile	Asn	Phe	Thr	Thr	Ala	Gly	
1				5				10						15		
GCC	ACT	GTG	CAA	AGC	TAC	ACA	AAC	TTT	ATC	AGA	GCT	GTT	CGC	GGT	CGT	96
Ala	Thr	Val	Gln	Ser	Tyr	Thr	Asn	Phe	Ile	Arg	Ala	Val	Arg	Gly	Arg	
			20					25						30		



- 63 -

TTA	ACA	ACT	GGA	GCT	GAT	GTG	AGA	CAT	GAA	ATA	CCA	GTG	TTG	CCA	AAC	144
Leu	Thr	Thr	Gly	Ala	Asp	Val	Arg	His	Glu	Ile	Pro	Val	Leu	Pro	Asn	
		35					40					45				
AGA	GTT	GGT	TTG	CCT	ATA	AAC	CAA	CGG	TTT	ATT	TTA	GTT	GAA	CTC	TCA	192
Arg	Val	Gly	Leu	Pro	Ile	Asn	Gln	Arg	Phe	Ile	Leu	Val	Glu	Leu	Ser	
	50					55					60					
AAT	CAT	GCA	GAG	CTT	TCT	GTT	ACA	TTA	GCG	CTG	GAT	GTC	ACC	AAT	GCA	240
Asn	His	Ala	Glu	Leu	Ser	Val	Thr	Leu	Ala	Leu	Asp	Val	Thr	Asn	Ala	
65					70					75					80	
TAT	GTG	GTA	GGC	TAC	CGT	GCT	GGA	AAT	AGC	GCA	TAT	TTC	TTT	CAT	CCT	288
Tyr	Val	Val	Gly	Tyr	Arg	Ala	Gly	Asn	Ser	Ala	Tyr	Phe	Phe	His	Pro	
				85					90					95		
GAC	AAT	CAG	GAA	GAT	GCA	GAA	GCA	ATC	ACT	CAT	CTT	TTC	ACT	GAT	GTT	336
Asp	Asn	Gln	Glu	Asp	Ala	Glu	Ala	Ile	Thr	His	Leu	Phe	Thr	Asp	Val	
			100					105					110			
CAA	AAT	CGA	TAT	ACA	TTC	GCC	TTT	GGT	GGT	AAT	TAT	GAT	AGA	CTT	GAA	384
Gln	Asn	Arg	Tyr	Thr	Phe	Ala	Phe	Gly	Gly	Asn	Tyr	Asp	Arg	Leu	Glu	
		115					120					125				
CAA	CTT	GCT	GGT	AAT	CTG	AGA	GAA	AAT	ATC	GAG	TTG	GGA	AAT	GGT	CCA	432
Gln	Leu	Ala	Gly	Asn	Leu	Arg	Glu	Asn	Ile	Glu	Leu	Gly	Asn	Gly	Pro	
	130					135					140					
CTA	GAG	GAG	GCT	ATC	TCA	GCG	CTT	TAT	TAT	TAC	AGT	ACT	GGT	GGC	ACT	480
Leu	Glu	Glu	Ala	Ile	Ser	Ala	Leu	Tyr	Tyr	Tyr	Ser	Thr	Gly	Gly	Thr	
145					150					155					160	
CAG	CTT	CCA	ACT	CTG	GCT	CGT	TCC	TTT	ATA	ATT	TGC	ATC	CAA	ATG	ATT	528
Gln	Leu	Pro	Thr	Leu	Ala	Arg	Ser	Phe	Ile	Ile	Cys	Ile	Gln	Met	Ile	
				165					170					175		
TCA	GAA	GCA	GCA	AGA	TTC	CAA	TAT	ATT	GAG	GGA	GAA	ATG	CGC	ACG	AGA	576
Ser	Glu	Ala	Ala	Arg	Phe	Gln	Tyr	Ile	Glu	Gly	Glu	Met	Arg	Thr	Arg	
			180					185					190			
ATT	AGG	TAC	AAC	CGG	AGA	TCT	GCA	CCA	GAT	CCT	AGC	GTA	ATT	ACA	CTT	624
Ile	Arg	Tyr	Asn	Arg	Arg	Ser	Ala	Pro	Asp	Pro	Ser	Val	Ile	Thr	Leu	
		195					200					205				
GAG	AAT	AGT	TGG	GGG	AGA	CTT	TCC	ACT	GCA	ATT	CAA	GAG	TCT	AAC	CAA	672
Glu	Asn	Ser	Trp	Gly	Arg	Leu	Ser	Thr	Ala	Ile	Gln	Glu	Ser	Asn	Gln	
	210					215					220					
GGA	GCC	TTT	GCT	AGT	CCA	ATT	CAA	CTG	CAA	AGA	CGT	AAT	GGT	TCC	AAA	720
Gly	Ala	Phe	Ala	Ser	Pro	Ile	Gln	Leu	Gln	Arg	Arg	Asn	Gly	Ser	Lys	
225					230					235					240	
TTC	AGT	GTG	TAC	GAT	GTG	AGT	ATA	TTA	ATC	CCT	ATC	ATA	GCT	CTC	ATG	768
Phe	Ser	Val	Tyr	Asp	Val	Ser	Ile	Leu	Ile	Pro	Ile	Ile	Ala	Leu	Met	
				245					250					255		

- 64 -

GTG TAT AGA TGC GCA CCT CCA CCA TCG TCA CAG TTT TAA  
 Val Tyr Arg Cys Ala Pro Pro Pro Ser Ser Gln Phe  
 260 265

807

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 268 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Ile Phe Pro Lys Gln Tyr Pro Ile Ile Asn Phe Thr Thr Ala Gly  
 1 5 10 15  
 Ala Thr Val Gln Ser Tyr Thr Asn Phe Ile Arg Ala Val Arg Gly Arg  
 20 25 30  
 Leu Thr Thr Gly Ala Asp Val Arg His Glu Ile Pro Val Leu Pro Asn  
 35 40 45  
 Arg Val Gly Leu Pro Ile Asn Gln Arg Phe Ile Leu Val Glu Leu Ser  
 50 55 60  
 Asn His Ala Glu Leu Ser Val Thr Leu Ala Leu Asp Val Thr Asn Ala  
 65 70 75 80  
 Tyr Val Val Gly Tyr Arg Ala Gly Asn Ser Ala Tyr Phe Phe His Pro  
 85 90 95  
 Asp Asn Gln Glu Asp Ala Glu Ala Ile Thr His Leu Phe Thr Asp Val  
 100 105 110  
 Gln Asn Arg Tyr Thr Phe Ala Phe Gly Gly Asn Tyr Asp Arg Leu Glu  
 115 120 125  
 Gln Leu Ala Gly Asn Leu Arg Glu Asn Ile Glu Leu Gly Asn Gly Pro  
 130 135 140  
 Leu Glu Glu Ala Ile Ser Ala Leu Tyr Tyr Tyr Ser Thr Gly Gly Thr  
 145 150 155 160  
 Gln Leu Pro Thr Leu Ala Arg Ser Phe Ile Ile Cys Ile Gln Met Ile  
 165 170 175  
 Ser Glu Ala Ala Arg Phe Gln Tyr Ile Glu Gly Glu Met Arg Thr Arg  
 180 185 190  
 Ile Arg Tyr Asn Arg Arg Ser Ala Pro Asp Pro Ser Val Ile Thr Leu  
 195 200 205

- 65 -

Glu Asn Ser Trp Gly Arg Leu Ser Thr Ala Ile Gln Glu Ser Asn Gln  
 210 215 220

Gly Ala Phe Ala Ser Pro Ile Gln Leu Gln Arg Arg Asn Gly Ser Lys  
 225 230 235 240

Phe Ser Val Tyr Asp Val Ser Ile Leu Ile Pro Ile Ile Ala Leu Met  
 245 250 255

Val Tyr Arg Cys Ala Pro Pro Pro Ser Ser Gln Phe  
 260 265

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1605 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1605
- (D) OTHER INFORMATION: /note= "product = "G-FIT"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AAG CTT ATG ATA TTC CCC AAA CAA TAC CCA ATT ATA AAC TTT ACC ACA	48
Lys Leu Met Ile Phe Pro Lys Gln Tyr Pro Ile Ile Asn Phe Thr Thr	
1 5 10 15	
GCG GGT GCC ACT GTG CAA AGC TAC ACA AAC TTT ATC AGA GCT GTT CGC	96
Ala Gly Ala Thr Val Gln Ser Tyr Thr Asn Phe Ile Arg Ala Val Arg	
20 25 30	
GGT CGT TTA ACA ACT GGA GCT GAT GTG AGA CAT GAA ATA CCA GTG TTG	144
Gly Arg Leu Thr Thr Gly Ala Asp Val Arg His Glu Ile Pro Val Leu	
35 40 45	
CCA AAC AGA GTT GGT TTG CCT ATA AAC CAA CGG TTT ATT TTA GTT GAA	192
Pro Asn Arg Val Gly Leu Pro Ile Asn Gln Arg Phe Ile Leu Val Glu	
50 55 60	
CTC TCA AAT CAT GCA GAG CTT TCT GTT ACA TTA GCG CTG GAT GTC ACC	240
Leu Ser Asn His Ala Glu Leu Ser Val Thr Leu Ala Leu Asp Val Thr	
65 70 75 80	
AAT GCA TAT GTG GTA GGC TAC CGT GCT GGA AAT AGC GCA TAT TTC TTT	288
Asn Ala Tyr Val Val Gly Tyr Arg Ala Gly Asn Ser Ala Tyr Phe Phe	
85 90 95	

- 66 -

CAT His	CCT Pro	GAC Asp	AAT Asn 100	CAG Gln	GAA Glu	GAT Asp	GCA Ala	GAA Glu 105	GCA Ala	ATC Ile	ACT Thr	CAT His	CTT Leu 110	TTC Phe	ACT Thr	336
GAT Asp	GTT Val	CAA Gln 115	AAT Asn	CGA Arg	TAT Tyr	ACA Thr	TTC Phe 120	GCC Ala	TTT Phe	GGT Gly	GGT Gly	AAT Asn 125	TAT Tyr	GAT Asp	AGA Arg	384
CTT Leu 130	GAA Glu	CAA Gln	CTT Leu	GCT Ala	GGT Gly	AAT Asn 135	CTG Leu	AGA Arg	GAA Glu	AAT Asn 140	ATC Ile	GAG Glu	TTG Leu	GGA Gly	AAT Asn	432
GGT Gly 145	CCA Pro	CTA Leu	GAG Glu	GAG Glu	GCT Ala 150	ATC Ile	TCA Ser	GCG Ala	CTT Leu	TAT Tyr 155	TAT Tyr	TAC Tyr	AGT Ser	ACT Thr	GGT Gly 160	480
GGC Gly	ACT Thr	CAG Gln	CTT Leu 165	CCA Pro	ACT Thr	CTG Leu	GCT Ala	CGT Arg 170	TCC Ser	TTT Phe	ATA Ile	ATT Ile	TGC Cys	ATC Ile 175	CAA Gln	528
ATG Met	ATT Ile	TCA Ser	GAA Glu 180	GCA Ala	GCA Ala	AGA Arg	TTC Phe	CAA Gln 185	TAT Tyr	ATT Ile	GAG Glu	GGA Gly	GAA Glu 190	ATG Met	CGC Arg	576
ACG Thr	AGA Arg	ATT Ile 195	AGG Arg	TAC Tyr	AAC Asn	CGG Arg	AGA Arg 200	TCT Ser	GCA Ala	CCA Pro	GAT Asp	CCT Pro 205	AGC Ser	GTA Val	ATT Ile	624
ACA Thr 210	CTT Leu	GAG Glu	AAT Asn	AGT Ser	TGG Trp	GGG Gly 215	AGA Arg	CTT Leu	TCC Ser	ACT Thr	GCA Ala 220	ATT Ile	CAA Gln	GAG Glu	TCT Ser	672
AAC Asn 225	CAA Gln	GGA Gly	GCC Ala	TTT Phe	GCT Ala	AGT Ser 230	CCA Pro	ATT Ile	CAA Gln	CTG Leu 235	CAA Gln	AGA Arg	CGT Arg	AAT Asn	GGT Gly 240	720
TCC Ser	AAA Lys	TTC Phe	AGT Ser	GTG Val 245	TAC Tyr	GAT Asp	GTG Val	AGT Ser	ATA Ile 250	TTA Leu	ATC Ile	CCT Pro	ATC Ile	ATA Ile 255	GCT Ala	768
CTC Leu	ATG Met	GTG Val	TAT Tyr 260	AGA Arg	TGC Cys	GCA Ala	CCT Pro	CCA Pro	CCA Pro	TCG Ser	TCA Ser	CAG Gln	TTT Phe 270	TCT Ser	CTT Leu	816
CTT Leu	ATA Ile	AGG Arg 275	CCA Pro	GTG Val	GTA Val	CCA Pro	AAT Asn 280	TTT Phe	AAT Asn	GCT Ala	GAT Asp	GTT Val 285	TGT Cys	ATG Met	GAT Asp	864
CCT Pro	GAG Glu	ATC Ile	CAA Gln	TTG Leu	GTG Val	CAG Gln 295	TCT Ser	GGA Gly	CCT Pro	GAG Glu	CTG Leu	AAG Lys	AAG Lys	CCT Pro	GGA Gly	912

- 67 -

GAG Glu 305	ACA Thr	GTC Val	AAG Lys	ATC Ile	TCC Ser 310	TGC Cys	AAG Lys	GCT Ala	TCT Ser	GGA Gly 315	TAT Tyr	ACC Thr	TTC Phe	GCA Ala	AAC Asn 320	960
TAT Tyr	GGA Gly	ATG Met	AAC Asn	TGG Trp 325	ATG Met	AAG Lys	CAG Gln	GCT Ala	CCA Pro 330	GGA Gly	AAG Lys	GGT Gly	TTA Leu	AAG Lys 335	TGG Trp	1008
ATG Met	GGC Gly	TGG Trp	ATA Ile 340	AAC Asn	ACC Thr	TAC Tyr	ACT Thr	GGA Gly 345	CAG Gln	TCA Ser	ACA Thr	TAT Tyr	GCT Ala 350	GAT Asp	GAC Asp	1056
TTC Phe	AAG Lys	GAA Glu 355	CGG Arg	TTT Phe	GCC Ala	TTC Phe	TCT Ser 360	TTG Leu	GAA Glu	ACC Thr	TCT Ser	GCC Ala 365	ACC Thr	ACT Thr	GCC Ala	1104
CAT His 370	TTG Leu	CAG Gln	ATC Ile	AAC Asn	AAC Asn 375	CTC Leu	AGA Arg	AAT Asn	GAG Glu	GAC Asp	TCG Ser 380	GCC Ala	ACA Thr	TAT Tyr	TTC Phe	1152
TGT Cys 385	GCA Ala	AGA Arg	CGA Arg	TTT Phe 390	GGG Gly	TTT Phe	GCT Ala	TAC Tyr	TGG Trp	GGC Gly 395	CAA Gln	GGG Gly	ACT Thr	CTG Leu	GTC Val 400	1200
AGT Ser	GTC Val	TCT Ser	GCA Ala	TCG Ser 405	ATA Ile	TCG Ser	AGC Ser	TCT Ser	GGT Gly 410	GGC Gly	GGT Gly	GGC Gly	TCG Ser	GGC Gly 415	GGT Gly	1248
GGT Gly	GGG Gly	TCG Ser	GGT Gly 420	GGC Gly	GGC Gly	GGA Gly	TCG Ser	GAT Asp 425	ATC Ile	CAG Gln	ATG Met	ACC Thr	CAG Gln 430	TCT Ser	CCA Pro	1296
TCC Ser	TCC Ser	TTA Leu 435	TCT Ser	GCC Ala	TCT Ser	CTG Leu	GGA Gly 440	GAA Glu	AGA Arg	GTC Val	AGT Ser	CTC Leu 445	ACT Thr	TGT Cys	CGG Arg	1344
GCA Ala 450	AGT Ser	CAG Gln	GAC Asp	ATT Ile	GGT Gly	AAT Asn 455	AGC Ser	TTA Leu	ACC Thr	TGG Trp	CTT Leu 460	TCA Ser	CAG Gln	GAA Glu	CCA Pro	1392
GAT Asp 465	GGA Gly	ACT Thr	ATT Ile	AAA Lys	CGC Arg 470	CTG Leu	ATC Ile	TAC Tyr	GCC Ala	ACA Thr 475	TCC Ser	AGT Ser	TTA Leu	GAT Asp 480	TCT Ser	1440
GGT Gly	GTC Val	CCC Pro	AAA Lys	AGG Arg 485	TTC Phe	AGT Ser	GGC Gly	AGT Ser	CGG Arg 490	TCT Ser	GGG Gly	TCA Ser	GAT Asp	TAT Tyr 495	TCT Ser	1488
CTC Leu	ACC Thr	ATC Ile	AGT Ser 500	AGC Ser	CTT Leu	GAG Glu	TCT Ser	GAA Glu 505	GAT Asp	TTT Phe	GTA Val	GTC Val	TAT Tyr 510	TAC Tyr	TGT Cys	1536
CTA Leu	CAA Gln	TAT Tyr 515	GCT Ala	ATT Ile	TTT Phe	CCG Pro	TAC Tyr 520	ACG Thr	TTC Phe	GGA Gly	GGG Gly	GGG Gly 525	ACC Thr	AAC Asn	CTG Leu	1584

- 68 -

GAA ATA AAA CGG GCT GAT TAA  
 Glu Ile Lys Arg Ala Asp  
 530 535

1605

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 534 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Lys Leu Met Ile Phe Pro Lys Gln Tyr Pro Ile Ile Asn Phe Thr Thr  
 1 5 10 15  
 Ala Gly Ala Thr Val Gln Ser Tyr Thr Asn Phe Ile Arg Ala Val Arg  
 20 25 30  
 Gly Arg Leu Thr Thr Gly Ala Asp Val Arg His Glu Ile Pro Val Leu  
 35 40 45  
 Pro Asn Arg Val Gly Leu Pro Ile Asn Gln Arg Phe Ile Leu Val Glu  
 50 55 60  
 Leu Ser Asn His Ala Glu Leu Ser Val Thr Leu Ala Leu Asp Val Thr  
 65 70 75 80  
 Asn Ala Tyr Val Val Gly Tyr Arg Ala Gly Asn Ser Ala Tyr Phe Phe  
 85 90 95  
 His Pro Asp Asn Gln Glu Asp Ala Glu Ala Ile Thr His Leu Phe Thr  
 100 105 110  
 Asp Val Gln Asn Arg Tyr Thr Phe Ala Phe Gly Gly Asn Tyr Asp Arg  
 115 120 125  
 Leu Glu Gln Leu Ala Gly Asn Leu Arg Glu Asn Ile Glu Leu Gly Asn  
 130 135 140  
 Gly Pro Leu Glu Glu Ala Ile Ser Ala Leu Tyr Tyr Tyr Ser Thr Gly  
 145 150 155 160  
 Gly Thr Gln Leu Pro Thr Leu Ala Arg Ser Phe Ile Ile Cys Ile Gln  
 165 170 175  
 Met Ile Ser Glu Ala Ala Arg Phe Gln Tyr Ile Glu Gly Glu Met Arg  
 180 185 190  
 Thr Arg Ile Arg Tyr Asn Arg Arg Ser Ala Pro Asp Pro Ser Val Ile  
 195 200 205

- 69 -

Thr Leu Glu Asn Ser Trp Gly Arg Leu Ser Thr Ala Ile Gln Glu Ser  
 210 215 220  
 Asn Gln Gly Ala Phe Ala Ser Pro Ile Gln Leu Gln Arg Arg Asn Gly  
 225 230 235 240  
 Ser Lys Phe Ser Val Tyr Asp Val Ser Ile Leu Ile Pro Ile Ile Ala  
 245 250 255  
 Leu Met Val Tyr Arg Cys Ala Pro Pro Pro Ser Ser Gln Phe Ser Leu  
 260 265 270  
 Leu Ile Arg Pro Val Val Pro Asn Phe Asn Ala Asp Val Cys Met Asp  
 275 280 285  
 Pro Glu Ile Gln Leu Val Gln Ser Gly Pro Glu Leu Lys Lys Pro Gly  
 290 295 300  
 Glu Thr Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Ala Asn  
 305 310 315 320  
 Tyr Gly Met Asn Trp Met Lys Gln Ala Pro Gly Lys Gly Leu Lys Trp  
 325 330 335  
 Met Gly Trp Ile Asn Thr Tyr Thr Gly Gln Ser Thr Tyr Ala Asp Asp  
 340 345 350  
 Phe Lys Glu Arg Phe Ala Phe Ser Leu Glu Thr Ser Ala Thr Thr Ala  
 355 360 365  
 His Leu Gln Ile Asn Asn Leu Arg Asn Glu Asp Ser Ala Thr Tyr Phe  
 370 375 380  
 Cys Ala Arg Arg Phe Gly Phe Ala Tyr Trp Gly Gln Gly Thr Leu Val  
 385 390 395 400  
 Ser Val Ser Ala Ser Ile Ser Ser Ser Gly Gly Gly Gly Ser Gly Gly  
 405 410 415  
 Gly Gly Ser Gly Gly Gly Gly Ser Asp Ile Gln Met Thr Gln Ser Pro  
 420 425 430  
 Ser Ser Leu Ser Ala Ser Leu Gly Glu Arg Val Ser Leu Thr Cys Arg  
 435 440 445  
 Ala Ser Gln Asp Ile Gly Asn Ser Leu Thr Trp Leu Ser Gln Glu Pro  
 450 455 460  
 Asp Gly Thr Ile Lys Arg Leu Ile Tyr Ala Thr Ser Ser Leu Asp Ser  
 465 470 475 480  
 Gly Val Pro Lys Arg Phe Ser Gly Ser Arg Ser Gly Ser Asp Tyr Ser  
 485 490 495

- 70 -

Leu Thr Ile Ser Ser Leu Glu Ser Glu Asp Phe Val Val Tyr Tyr Cys  
                   500                                  505                                  510

Leu Gln Tyr Ala Ile Phe Pro Tyr Thr Phe Gly Gly Gly Thr Asn Leu  
                   515                                  520                                  525

Glu Ile Lys Arg Ala Asp  
           530

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..45
- (D) OTHER INFORMATION: /note= "product = "new linker/"  
                           info: new linker"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TCG	AGC	TCC	TCC	GGA	TCT	TCA	TCT	AGC	GGT	TCC	AGC	TCG	AGT	GGA	45
Ser	Ser	Ser	Ser	Gly	Ser	Ser	Ser	Ser	Gly	Ser	Ser	Ser	Ser	Gly	
1				5					10					15	

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Ser	Ser	Ser	Ser	Gly	Ser	Ser	Ser	Ser	Gly	Ser	Ser	Ser	Ser	Gly
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear



- 71 -

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..45

(D) OTHER INFORMATION: /note= "product = "old linker/  
protein info: old linker"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGA	GGA	GGA	GGA	TCT	GGA	GGA	GGA	GGA	TCT	GGA	GGA	GGA	GGA	TCT	45
Gly	Gly	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Ser	
1				5					10					15	

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gly	Gly	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Ser	Gly	Gly	Gly	Gly	Ser
1				5					10					15

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2001 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..2001

(D) OTHER INFORMATION: /note= "product = "741sFv-PE40"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAT	CCT	GAG	ATC	CAA	TTG	GTG	CAG	TCT	GGA	CCT	GAG	CTG	AAG	AAG	CCT	48
Asp	Pro	Glu	Ile	Gln	Leu	Val	Gln	Ser	Gly	Pro	Glu	Leu	Lys	Lys	Pro	
1				5					10						15	

GGA	GAG	ACA	GTC	AAG	ATC	TCC	TGC	AAG	GCT	TCT	GGG	TAT	ACC	TTC	ACA	96
Gly	Glu	Thr	Val	Lys	Ile	Ser	Cys	Lys	Ala	Ser	Gly	Tyr	Thr	Phe	Thr	
			20						25						30	

- 72 -

AAC TAT GGA ATG AAC TGG GTG AAG CAG GCT CCA GGA AAG GGT TTA AAG Asn Tyr Gly Met Asn Trp Val Lys Gln Ala Pro Gly Lys Gly Leu Lys 35 40 45	144
TGG ATG GGC TGG ATA AAC ACC AAC ACT GGA GAG CCA ACA TAT GCT GAA Trp Met Gly Trp Ile Asn Thr Asn Thr Gly Glu Pro Thr Tyr Ala Glu 50 55 60	192
GAG TTC AAG GGA CGG TTT GCC TTC TCT TTG GAA ACC TCT GCC AGC ACT Glu Phe Lys Gly Arg Phe Ala Phe Ser Leu Glu Thr Ser Ala Ser Thr 65 70 75 80	240
GCC TAT TTG CAG ATC AAC AAC CTC AAA AAT GAG GAC ACG GCT ACA TAT Ala Tyr Leu Gln Ile Asn Asn Leu Lys Asn Glu Asp Thr Ala Thr Tyr 85 90 95	288
TTC TGT GGA AGG CAA TTT ATT ACC TAC GGC GGG TTT GCT AAC TGG GGC Phe Cys Gly Arg Gln Phe Ile Thr Tyr Gly Gly Phe Ala Asn Trp Gly 100 105 110	336
CAA GGG ACT CTG GTC ACT GTC TCT GCA TCG AGC TCC TCC GGA TCT TCA Gln Gly Thr Leu Val Thr Val Ser Ala Ser Ser Ser Ser Gly Ser Ser 115 120 125	384
TCT AGC GGT TCC AGC TCG AGC GAT ATC GTC ATG ACC CAG TCT CCT AAA Ser Ser Gly Ser Ser Ser Ser Asp Ile Val Met Thr Gln Ser Pro Lys 130 135 140	432
TTC ATG TCC ACG TCA GTG GGA GAC AGG GTC AGC ATC TCC TGC AAG GCC Phe Met Ser Thr Ser Val Gly Asp Arg Val Ser Ile Ser Cys Lys Ala 145 150 155 160	480
AGT CAG GAT GTG AGT ACT GCT GTA GCC TGG TAT CAA CAA AAA CCA GGG Ser Gln Asp Val Ser Thr Ala Val Ala Trp Tyr Gln Gln Lys Pro Gly 165 170 175	528
CAA TCT CCT AAA CTA CTG ATT TAC TGG ACA TCC ACC CGG CAC ACT GGA Gln Ser Pro Lys Leu Leu Ile Tyr Trp Thr Ser Thr Arg His Thr Gly 180 185 190	576
GTC CCT GAT CCG TTC ACA GGC AGT GGA TCT GGG ACA GAT TAT ACT CTC Val Pro Asp Pro Phe Thr Gly Ser Gly Ser Gly Thr Asp Tyr Thr Leu 195 200 205	624
ACC ATC AGC AGT GTG CAG GCT GAA GAC CTG GCA CTT CAT TAC TGT CAG Thr Ile Ser Ser Val Gln Ala Glu Asp Leu Ala Leu His Tyr Cys Gln 210 215 220	672
CAA CAT TAT AGA GTG GCC TAC ACG TTC GGA AGG GGG ACC AAG CTG GAG Gln His Tyr Arg Val Ala Tyr Thr Phe Gly Arg Gly Thr Lys Leu Glu 225 230 235 240	720
ATA AAA CGG GCT GAT GCT GCA CCA ACT GTA TCC ATC TTC CCA CCA TCC Ile Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser 245 250 255	768

- 73 -

AGT Ser	GAG Glu	CAG Gln	TTT Phe 260	GAG Glu	GGC Gly	GGC Gly	AGC Ser	CTG Leu	GCC Ala	GCG Ala	CTG Leu	AAC Asn	GCG Ala	CAC His	CAG Gln	816
GCT Ala	TGC Cys	CAC His 275	CTG Leu	CCG Pro	CTG Leu	GAG Glu	ACT Thr 280	TTC Phe	ACC Thr	CGT Arg	CAT His	CGC Arg 285	CAG Gln	CCG Pro	CGC Arg	864
GGC Gly	TGG Trp 290	GAA Glu	CAA Gln	CTG Leu	GAG Glu	CAG Gln 295	TGC Cys	GGC Gly	TAT Tyr	CCG Pro	GTG Val 300	CAG Gln	CGG Arg	CTG Leu	GTC Val	912
GCC Ala 305	CTC Leu	TAC Tyr	CTG Leu	GCG Ala	GCG Ala	CGG Arg 310	CTG Leu	TCG Ser	TGG Trp	AAC Asn 315	CAG Gln	GTC Val	GAC Asp	CAG Gln	GTC Val 320	960
ATC Ile	CGC Arg	AAC Asn	GCC Ala 325	CTG Leu	GCC Ala	AGC Ser	CCC Pro	GGC Gly	AGC Ser 330	GGC Gly	GAC Gly	CTG Asp	GGC Leu	GAA Gly	GAA Glu 335	1008
GCG Ala	ATC Ile	CGC Arg	GAG Glu 340	CAG Gln	CCG Pro	GAG Glu	CAG Gln 345	GCC Ala	CGT Arg	CTG Leu	GCC Ala	CTG Leu	ACC Thr 350	CTG Leu	GCC Ala	1056
GCC Ala	GCC Ala	GAG Glu 355	AGC Ser	GAG Glu	CGC Arg	TTC Phe	GTC Val 360	CGG Arg	CAG Gln	GGC Gly	ACC Thr	GGC Gly	AAC Asn	GAC Asp	GAG Glu	1104
GCC Ala 370	GGC Gly	GCG Ala	GCC Ala	AAC Asn	GCC Ala	GAC Asp 375	GTG Val	GTG Val	AGC Ser	CTG Leu	ACC Thr 380	TGC Cys	CCG Pro	GTC Val	GCC Ala	1152
GCC Ala 385	GGT Gly	GAA Glu	TGC Cys	GCG Ala	GGC Gly	CCG Pro 390	GCG Ala	GAC Asp	AGC Ser	GGC Gly 395	GAC Asp	GCC Ala	CTG Leu	CTG Leu	GAG Glu 400	1200
CGC Arg	AAC Asn	TAT Tyr	CCC Pro	ACT Thr 405	GGC Gly	GCG Ala	GAG Glu	TTC Phe	CTC Leu 410	GGC Gly	GAC Asp	GGC Gly	GGC Gly	GAC Asp	GTC Val 415	1248
AGC Ser	TTC Phe	AGC Ser	AAC Asn 420	CGC Arg	GGC Gly	ACG Thr	CAG Gln	AAC Asn 425	TGG Trp	ACG Thr	GTG Val	GAG Glu	CGG Arg 430	CTG Leu	CTC Leu	1296
CAG Gln	GCG Ala	CAC His 435	CGC Arg	CAA Gln	CTG Leu	GAG Glu	GAG Glu 440	CGC Arg	GGC Gly	TAT Tyr	GTG Val	TTC Phe 445	GTC Val	GGC Gly	TAC Tyr	1344
CAC His 450	GGC Gly	ACC Thr	TTC Phe	CTC Leu	GAA Glu	GCG Ala 455	GCG Ala	CAA Gln	AGC Ser	ATC Ile	GTG Val 460	TTC Phe	GGC Gly	GGG Gly	GTC Val	1392
CGC Arg 465	GCG Ala	CGC Arg	AGC Ser	CAG Gln	GAC Asp 470	CTC Leu	GAC Asp	GCG Ala	ATC Ile	TGG Trp 475	CGC Arg	GGT Gly	TTC Phe	TAT Tyr	ATC Ile 480	1440

- 74 -

GCC GGC GAT CCG GCG CTG GCC TAC GGC TAC GCC CAG GAC CAG GAA CCC	1488
Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr Ala Gln Asp Gln Glu Pro	
485 490 495	
GAC GCA CGC GGC CGG ATC CGC AAC GGT GCC CTG CTG CGG GTC TAT GTG	1536
Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala Leu Leu Arg Val Tyr Val	
500 505 510	
CCG CGC TCG AGC CTG CCG GGC TTC TAC CGC ACC AGC CTG ACC CTG GCC	1584
Pro Arg Ser Ser Leu Pro Gly Phe Tyr Arg Thr Ser Leu Thr Leu Ala	
515 520 525	
GCG CCG GAG GCG GCG GGC GAG GTC GAA CGG CTG ATC GGC CAT CCG CTG	1632
Ala Pro Glu Ala Ala Gly Glu Val Glu Arg Leu Ile Gly His Pro Leu	
530 535 540	
CCG CTG CGC CTG GAC GCC ATC ACC GGC CCC GAG GAG GAA GGC GGG CGC	1680
Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro Glu Glu Glu Gly Gly Arg	
545 550 555 560	
CTG GAG ACC ATT CTC GGC TGG CCG CTG GCC GAG CGC ACC GTG GTG ATT	1728
Leu Glu Thr Ile Leu Gly Trp Pro Leu Ala Glu Arg Thr Val Val Ile	
565 570 575	
CCC TCG GCG ATC CCC ACC GAC CCG CGC AAC GTC GGC GGC GAC CTC GAC	1776
Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val Gly Gly Asp Leu Asp	
580 585 590	
CCG TCC AGC ATC CCC GAC AAG GAA CAG GCG ATC AGC GCC CTG CCG GAC	1824
Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu Pro Asp	
595 600 605	
TAC GCC AGC CAG CCC GGC AAA CCG CCG CGC GAG GAC CTG AAG TAA CTG	1872
Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Leu Lys * Leu	
610 615 620	
CCG CGA CCG GCC GGC TCC CTT CGC AGG AGC CGG CCT TCT CGG GGC CTG	1920
Pro Arg Pro Ala Gly Ser Leu Arg Arg Ser Arg Pro Ser Arg Gly Leu	
625 630 635 640	
GCC ATA CAT CAG GTT TTC CTG ATG CCA GCC CAA TCG AAT ATG AAT TGA	1968
Ala Ile His Gln Val Phe Leu Met Pro Ala Gln Ser Asn Met Asn *	
645 650 655	
TCC TCT AGA GTC GAC CTG CAG GCA TGC AAG CTT	2001
Ser Ser Arg Val Asp Leu Gln Ala Cys Lys Leu	
660 665	

## (2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 667 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- 75 -

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```

Asp Pro Glu Ile Gln Leu Val Gln Ser Gly Pro Glu Leu Lys Lys Pro
 1           5           10           15
Gly Glu Thr Val Lys Ile Ser Cys Lys Ala Ser Gly Tyr Thr Phe Thr
          20           25           30
Asn Tyr Gly Met Asn Trp Val Lys Gln Ala Pro Gly Lys Gly Leu Lys
          35           40           45
Trp Met Gly Trp Ile Asn Thr Asn Thr Gly Glu Pro Thr Tyr Ala Glu
          50           55           60
Glu Phe Lys Gly Arg Phe Ala Phe Ser Leu Glu Thr Ser Ala Ser Thr
          65           70           75           80
Ala Tyr Leu Gln Ile Asn Asn Leu Lys Asn Glu Asp Thr Ala Thr Tyr
          85           90           95
Phe Cys Gly Arg Gln Phe Ile Thr Tyr Gly Gly Phe Ala Asn Trp Gly
          100          105          110
Gln Gly Thr Leu Val Thr Val Ser Ala Ser Ser Ser Ser Gly Ser Ser
          115          120          125
Ser Ser Gly Ser Ser Ser Ser Asp Ile Val Met Thr Gln Ser Pro Lys
          130          135          140
Phe Met Ser Thr Ser Val Gly Asp Arg Val Ser Ile Ser Cys Lys Ala
          145          150          155          160
Ser Gln Asp Val Ser Thr Ala Val Ala Trp Tyr Gln Gln Lys Pro Gly
          165          170          175
Gln Ser Pro Lys Leu Leu Ile Tyr Trp Thr Ser Thr Arg His Thr Gly
          180          185          190
Val Pro Asp Pro Phe Thr Gly Ser Gly Ser Gly Thr Asp Tyr Thr Leu
          195          200          205
Thr Ile Ser Ser Val Gln Ala Glu Asp Leu Ala Leu His Tyr Cys Gln
          210          215          220
Gln His Tyr Arg Val Ala Tyr Thr Phe Gly Arg Gly Thr Lys Leu Glu
          225          230          235          240
Ile Lys Arg Ala Asp Ala Ala Pro Thr Val Ser Ile Phe Pro Pro Ser
          245          250          255
Ser Glu Gln Phe Glu Gly Gly Ser Leu Ala Ala Leu Asn Ala His Gln
          260          265          270

```

- 76 -

Ala Cys His Leu Pro Leu Glu Thr Phe Thr Arg His Arg Gln Pro Arg  
 275 280 285  
 Gly Trp Glu Gln Leu Glu Gln Cys Gly Tyr Pro Val Gln Arg Leu Val  
 290 295 300  
 Ala Leu Tyr Leu Ala Ala Arg Leu Ser Trp Asn Gln Val Asp Gln Val  
 305 310 315 320  
 Ile Arg Asn Ala Leu Ala Ser Pro Gly Ser Gly Gly Asp Leu Gly Glu  
 325 330 335  
 Ala Ile Arg Glu Gln Pro Glu Gln Ala Arg Leu Ala Leu Thr Leu Ala  
 340 345 350  
 Ala Ala Glu Ser Glu Arg Phe Val Arg Gln Gly Thr Gly Asn Asp Glu  
 355 360 365  
 Ala Gly Ala Ala Asn Ala Asp Val Val Ser Leu Thr Cys Pro Val Ala  
 370 375 380  
 Ala Gly Glu Cys Ala Gly Pro Ala Asp Ser Gly Asp Ala Leu Leu Glu  
 385 390 395 400  
 Arg Asn Tyr Pro Thr Gly Ala Glu Phe Leu Gly Asp Gly Gly Asp Val  
 405 410 415  
 Ser Phe Ser Asn Arg Gly Thr Gln Asn Trp Thr Val Glu Arg Leu Leu  
 420 425 430  
 Gln Ala His Arg Gln Leu Glu Glu Arg Gly Tyr Val Phe Val Gly Tyr  
 435 440 445  
 His Gly Thr Phe Leu Glu Ala Ala Gln Ser Ile Val Phe Gly Gly Val  
 450 455 460  
 Arg Ala Arg Ser Gln Asp Leu Asp Ala Ile Trp Arg Gly Phe Tyr Ile  
 465 470 475 480  
 Ala Gly Asp Pro Ala Leu Ala Tyr Gly Tyr Ala Gln Asp Gln Glu Pro  
 485 490 495  
 Asp Ala Arg Gly Arg Ile Arg Asn Gly Ala Leu Leu Arg Val Tyr Val  
 500 505 510  
 Pro Arg Ser Ser Leu Pro Gly Phe Tyr Arg Thr Ser Leu Thr Leu Ala  
 515 520 525  
 Ala Pro Glu Ala Ala Gly Glu Val Glu Arg Leu Ile Gly His Pro Leu  
 530 535 540  
 Pro Leu Arg Leu Asp Ala Ile Thr Gly Pro Glu Glu Glu Gly Gly Arg  
 545 550 555 560

- 77 -

Leu Glu Thr Ile Leu Gly Trp Pro Leu Ala Glu Arg Thr Val Val Ile  
 565 570 575  
 Pro Ser Ala Ile Pro Thr Asp Pro Arg Asn Val Gly Gly Asp Leu Asp  
 580 585 590  
 Pro Ser Ser Ile Pro Asp Lys Glu Gln Ala Ile Ser Ala Leu Pro Asp  
 595 600 605  
 Tyr Ala Ser Gln Pro Gly Lys Pro Pro Arg Glu Asp Leu Lys \* Leu  
 610 615 620  
 Pro Arg Pro Ala Gly Ser Leu Arg Arg Ser Arg Pro Ser Arg Gly Leu  
 625 630 635 640  
 Ala Ile His Gln Val Phe Leu Met Pro Ala Gln Ser Asn Met Asn \*  
 645 650 655  
 Ser Ser Arg Val Asp Leu Gln Ala Cys Lys Leu  
 660 665

- 78 -

CLAIMS

1 1. A single-chain Fv (sFv) polypeptide defining a  
2 binding site which exhibits the immunological binding  
3 properties of an immunoglobulin molecule which binds  
4 c-erbB-2 or a c-erbB-2-related tumor antigen, said sFv  
5 comprising at least two polypeptide domains connected  
6 by a polypeptide linker spanning the distance between  
7 the C-terminus of one domain and the N-terminus of the  
8 other, the amino acid sequence of each of said  
9 polypeptide domains comprising a set of complementarity  
10 determining regions (CDRs) interposed between a set of  
11 framework regions (FRs), said CDRs conferring  
12 immunological binding to said c-erbB-2 or c-erbB-2-  
13 related tumor antigen.

1 2. The single-chain Fv polypeptide of claim 1  
2 wherein said CDRs are substantially homologous with the  
3 CDRs of the c-erbB-2-binding immunoglobulin molecules  
4 selected from the group consisting of 520C9, 741F8, and  
5 454C11 monoclonal antibodies.

1 3. The single-chain Fv polypeptide of claim 2  
2 wherein the amino acid sequence of each of said sFv  
3 CDRs and each of said FRs are substantially homologous  
4 with the amino acid sequence of CDRs and FRs of the  
5 variable region of 520C9 antibody.

1 4. The single-chain Fv polypeptide of claim 1  
2 wherein said polypeptide linker comprises the amino  
3 acid sequence as set forth in the Sequence Listing as  
4 amino acid residue numbers 118 through 133 in SEQ ID  
5 NO:4.



- 79 -

1 5. The single-chain Fv polypeptide of claim 1  
2 wherein said polypeptide linker comprises an amino acid  
3 sequence selected from the group of sequences set forth  
4 as amino acid residues 116-135 in SEQ ID NO:6, or 122-  
5 135 in SEQ. ID NO:15 and the amino acid sequences set  
6 forth in SEQ ID NO: 12 and SEQ ID NO: 14.

1 6. The single-chain Fv polypeptide of claim 1  
2 further comprising a remotely detectable moiety bound  
3 thereto to permit imaging of a cell bearing said  
4 c-erbB-2-related tumor antigen.

1 7. The single-chain Fv polypeptide of claim 6  
2 wherein said remotely detectable moiety comprises a  
3 radioactive atom.

1 8. The single-chain Fv polypeptide of claim 1  
2 further comprising, linked to the N or C terminus of  
3 said linked domains, a third polypeptide domain  
4 comprising an amino acid sequence defining CDRs  
5 interposed between FRs and defining a second  
6 immunologically active site.

1 9. The single-chain Fv polypeptide of claim 8,  
2 further comprising a fourth polypeptide domain, wherein  
3 said third and fourth polypeptide domains together  
4 comprise a second site which immunologically binds a  
5 c-erbB-2-related tumor antigen.

1 10. The single-chain Fv polypeptide of claim 1 or 7  
2 further comprising a toxin linked to the N or C  
3 terminus of said linked domain.

- 80 -

1 11. The single-chain Fv polypeptide of claim 10  
2 wherein said toxin comprises a toxic portion selected  
3 from the group: Pseudomonas exotoxin, ricin, ricin A  
4 chain, phytolectin and diphtheria toxin.

1 12. The single-chain Fv polypeptide of claim 10  
2 wherein said toxin comprises at least a portion of the  
3 ricin A chain.

1 13. A DNA sequence encoding the polypeptide chain of  
2 claim 1.

1 14. A method of producing a single chain polypeptide  
2 having specificity for a c-erbB-2-related tumor  
3 antigen, said method comprising the steps of:  
4 (a) transfecting the DNA of claim 13 into a  
5 host cell to produce a transformant; and  
6 (b) culturing said transformant to produce  
7 said single-chain polypeptide.

1 15. A method of imaging a tumor expressing a  
2 c-erbB-2-related antigen, said method comprising the  
3 steps of:  
4 (a) providing an imaging agent comprising the  
5 polypeptide of claim 7;  
6 (b) administering to a mammal harboring said  
7 tumor an amount of said imaging agent together with a  
8 physiologically-acceptable carrier sufficient to permit  
9 extracorporeal detection of said tumor after allowing  
10 said agent to bind to said tumor; and  
11 (c) detecting the location of said remotely  
12 detectable moiety in said subject to obtain an image of  
13 said tumor.

- 81 -

1 16. A host cell transfected with a DNA of claim 13.

1 17. A method of inhibiting in vivo growth of a tumor  
2 expressing a c-erbB-2-related antigen, said method  
3 comprising:

4 administering to a patient harboring the tumor a  
5 tumor inhibiting amount of a therapeutic agent  
6 comprising a single-chain Fv of claim 1 and at least a  
7 first moiety peptide bonded thereto, said first moiety  
8 having the ability to limit the proliferation of a  
9 tumor cell.

1 18. The method of claim 17 wherein said first moiety  
2 comprises a cell toxin or a toxic fragment thereof.

1 19. The method of claim 17 wherein said first moiety  
2 comprises a radioisotope sufficiently radioactive to  
3 inhibit proliferation of said tumor cell.

1 20. A DNA sequence encoding the polypeptide chain of  
2 claim 10.

1/5

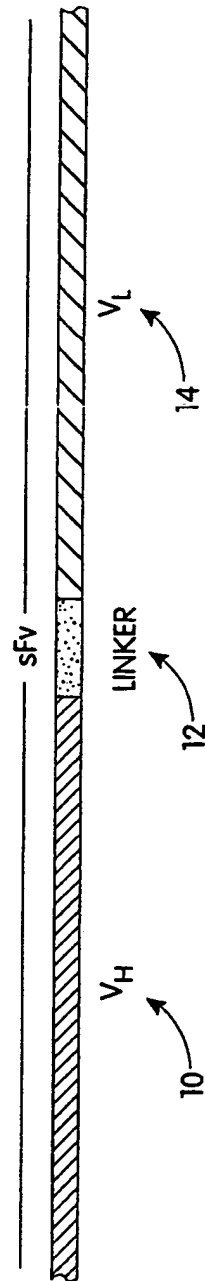


Fig. 1A

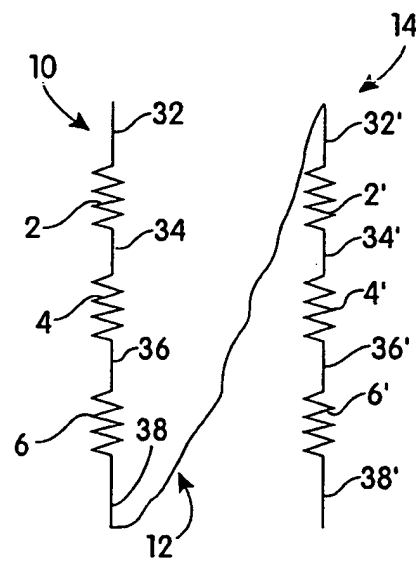


Fig. 1B

3/5

Fig. 2A

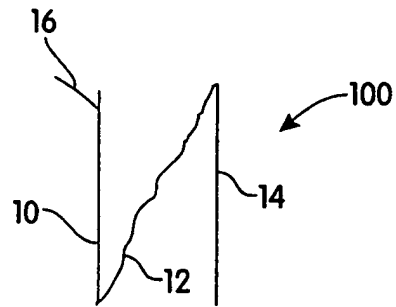


Fig. 2B

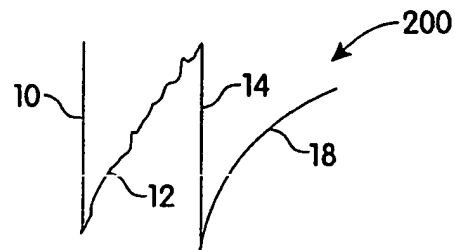


Fig. 2C

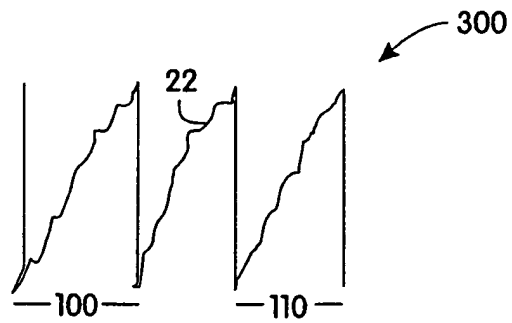


Fig. 2D

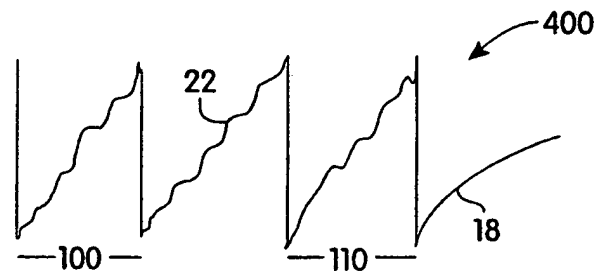
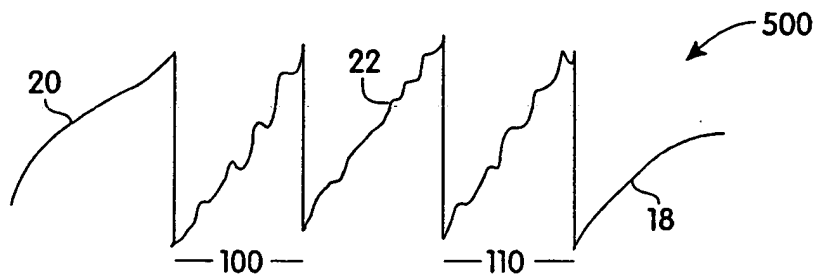


Fig. 2E



4/5

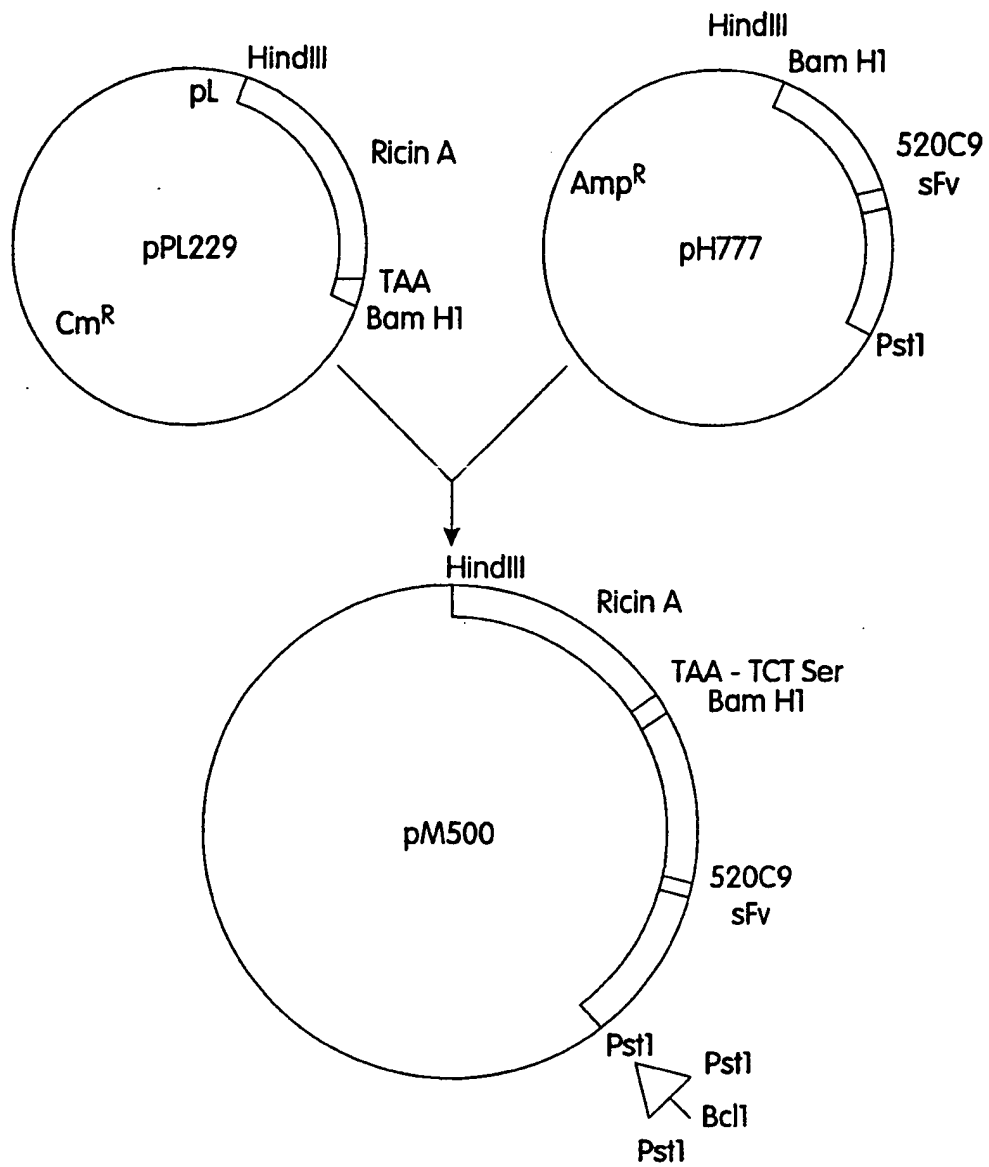


Fig. 3

5/5

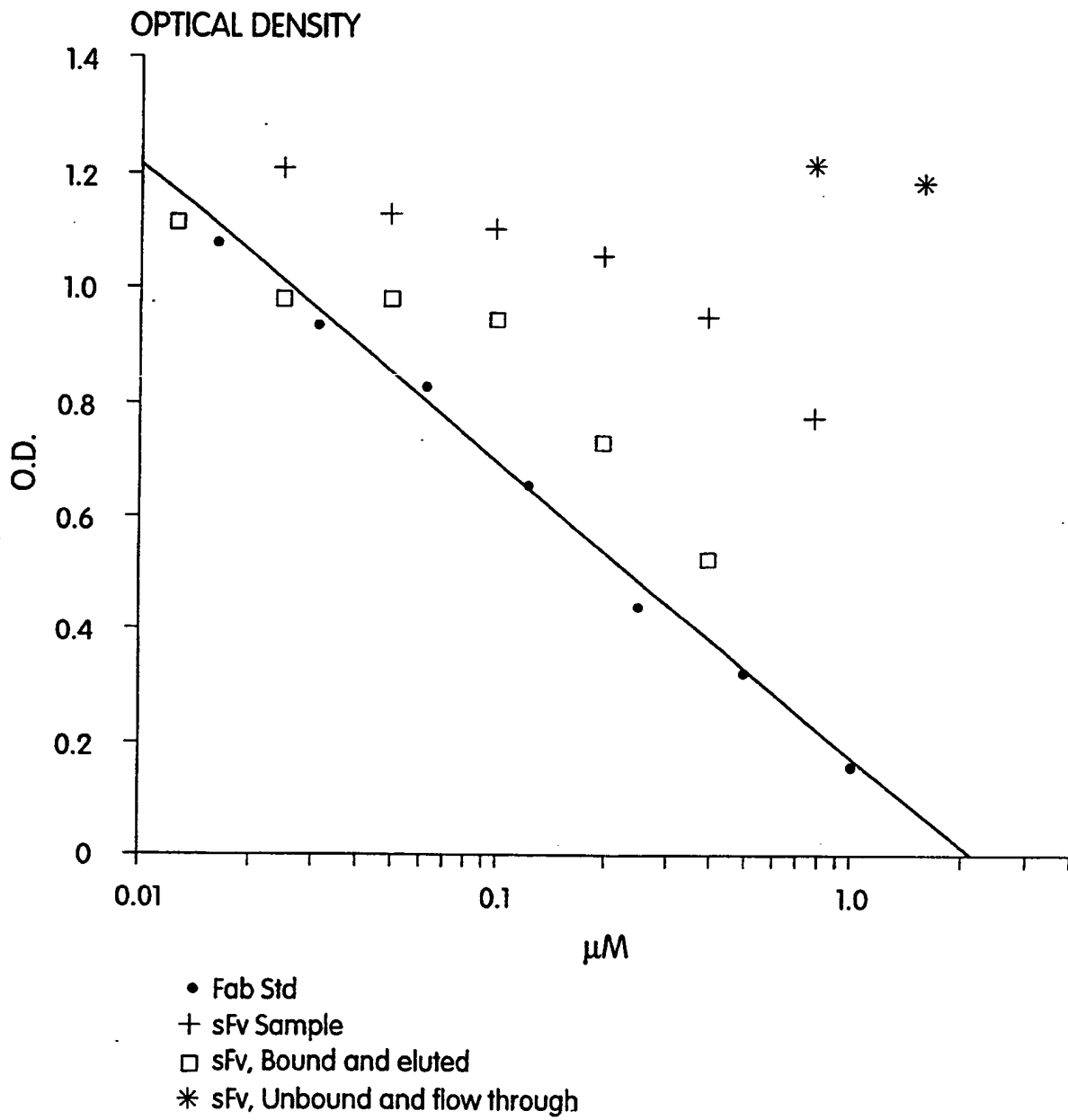


Fig. 4